



UNIL | Université de Lausanne



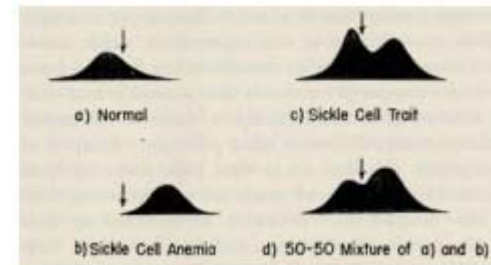
Genes: from sequence to function

**We got the sequence: now what? (part 2)**

# Position-independent gene identification

## [1. Starting from the protein product]

- It is mostly a pre-genomic strategy, relying on protein information and on biochemical notions
- The most famous example is Sickle Cell Anemia, where hemoglobin was shown to be different in patients vs. controls (1949)



- Obviously, the genetic defect should lie in the DNA encoding for the globins

Database - Windows Internet Explorer

http://www.ncbi.nlm.nih.gov/Database/index.html

File Edit View Favorites Tools Help

Database

NCBI **Databases**

PubMed Entrez BLAST OMIM Books TaxBrowser Structure

Search Entrez for Go

Entrez  
NCBI Web Site

NCBI

Site Map  
Guide to  
resources

Entrez H  
Help docu  
for the En

Entrez T  
Conserved Domains  
3D Domains  
Gene

Entrez C  
Query  
Search a  
Entrez da

Entrez T  
HomoloGene  
Links to a  
Entrez to

Batch En  
Utilities  
NCBI H

In-depth  
NCBI res

LinkOut  
External links to  
related resources

Literature  
Databases  
PubMed, OMIM,  
Books, PubMed  
Central

Related  
Resources

BLAST

the integrated, text-based search and retrieval system used for the major databases, including PubMed, Nucleotide and sequences, Protein Structures, Complete Genomes, and others. Click on the graphic below for a more view of Entrez integration.

The complete list of Entrez databases can be viewed in the search pull down menu.

**Nucleotide Databases**

<a href="#">dbEST</a>	<a href="#">PopSet</a>
<a href="#">dbGSS</a>	<a href="#">Probe</a>
<a href="#">dbSNP</a>	<a href="#">RefSeq</a>
<a href="#">dbSTS</a>	<a href="#">SRA</a>
<a href="#">Nucleotide</a>	<a href="#">TPA</a>
<a href="#">GenBank</a>	<a href="#">Trace Archive</a>
<a href="#">HomoloGene</a>	<a href="#">UniGene</a>
<a href="#">MGC</a>	<a href="#">UniSTS</a>

start

splicing-biochemi...

Eudora - [In]

2009-2010

Carlo\_Rivolta\_M...

Carlo\_Rivolta\_Ma...

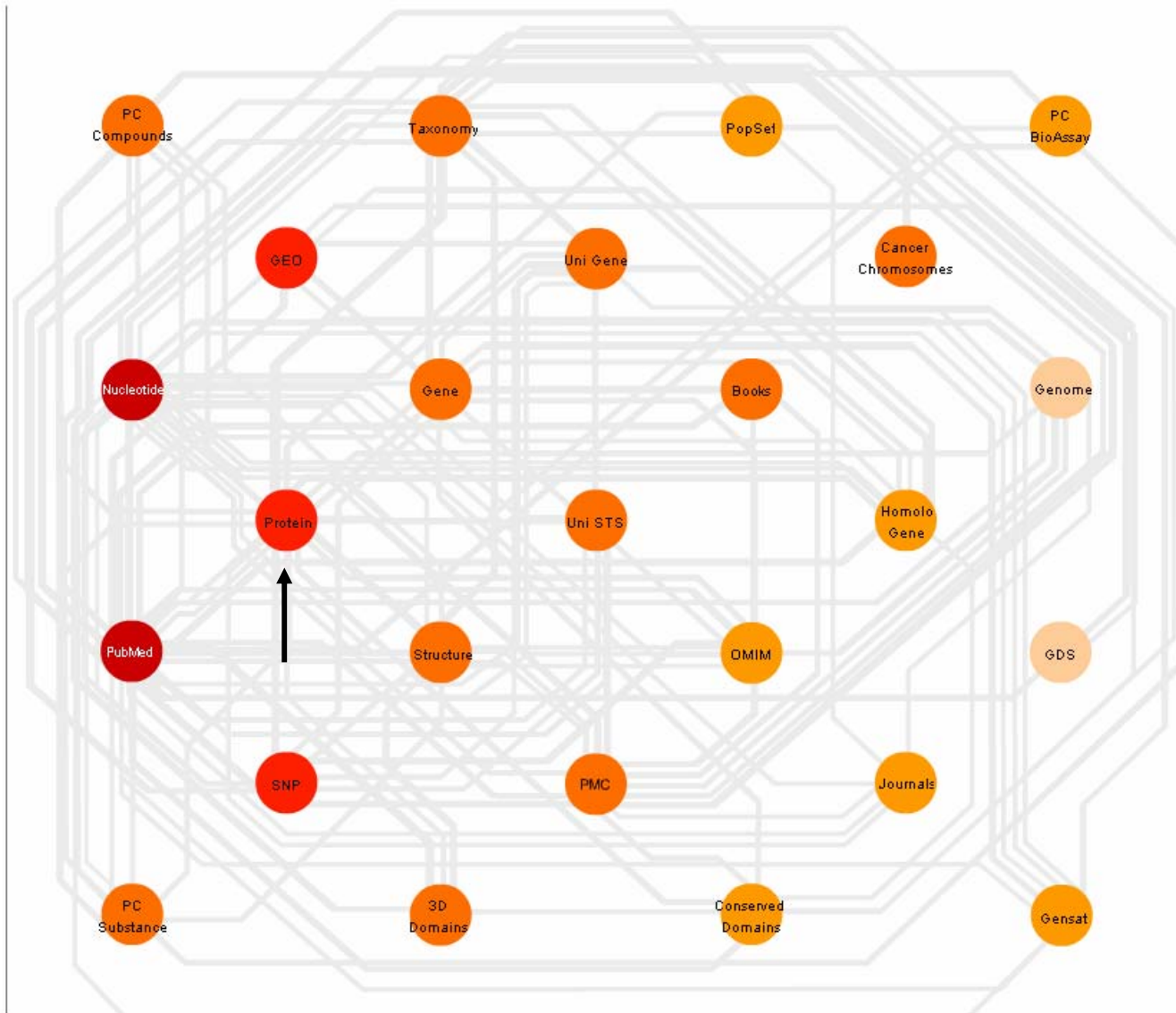
Database - Wind...

EditPad Lite

Internet

100%

10:42



Database - Windows Internet Explorer

http://www.ncbi.nlm.nih.gov/Database/index.html

File Edit View Favorites Tools Help

Database

NCBI Databases

PubMed Entrez BLAST OMIM Books TaxBrowser Structure

Search Entrez for Go

Entrez  
NCBI Web Site

NCBI

Site Map  
Guide to  
resource

Entrez H  
Help docu  
for the En

Entrez T  
Conserved Domains  
3D Domains  
Gene

Entrez G  
Genome Project  
dbGaP

Query  
Search a  
Entrez da

Entrez T  
HomoloGene  
Links to a  
Journals

Entrez to  
MeSH

Batch Ent  
NLM Catalog

Utilities  
OMIA  
OMIM

NCBI Ha  
PMC

In-depth  
PopSet

NCBI res  
Probe

Protein Clusters

The integrated, text-based search and retrieval system used for the major databases, including PubMed, Nucleotide and sequences, Protein Structures, Complete Genomes, and others. Click on the graphic below for a more view of Entrez integration.

The complete list of Entrez databases can be viewed in the search pull down menu.

Nucleotide Databases

smn - Protein Results - Windows Internet Explorer

http://www.ncbi.nlm.nih.gov/sites/entrez

File Edit View Favorites Tools Help

smn - Protein Results

NCBI Protein

All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books

Search Protein for smn Go Clear Save Search

Limits Preview/Index History Board Details

Display Summary Show 20 Sort By Send to

All: 3605 Bacteria: 12 RefSeq: 1523 Related Structures: 2031

This search in Gene shows 1283 results, including:

- [Smn](#) (*Drosophila melanogaster*): survival motor neuron
- [SMN](#) (*Bos taurus*): survival motor neuron
- [SMN](#) (*Gallus gallus*): survival motor neuron

Items 1 - 20 of 3605 Page 1 of 181 Next

1: [CAA45273](#) Reports Conserved Domains, BLink, Links  
SmN [Mus musculus]  
gi|407470|emb|CAA45273.1|[407470]

2: [P63163](#) Reports Conserved Domains, BLink, Links  
RecName: Full=Small nuclear ribonucleoprotein-associated protein N; Short=snRNP-N; Short=Sm protein N; Short=Sm-N;  
Short=SmN; Short=Sm-D; AltName: Full=Tissue-specific-splicing protein  
gi|52783795|sp|P63163.1|RSMN\_MOUSE[52783795]

3: [P63162](#) Reports Conserved Domains, BLink, Links  
RecName: Full=Small nuclear ribonucleoprotein-associated protein N; Short=snRNP-N; Short=Sm protein N; Short=Sm-N;  
Short=SmN; Short=Sm-D; AltName: Full=Tissue-specific-splicing protein  
gi|52783794|sp|P63162.1|RSMN\_HUMAN[52783794]

4: [P63164](#) Reports Conserved Domains, BLink, Links  
RecName: Full=Small nuclear ribonucleoprotein-associated protein N; Short=snRNP-N; Short=Sm protein N; Short=Sm-N;  
Short=SmN; Short=Sm-D  
gi|52783793|sp|P63164.1|RSMN\_RAT[52783793]

5: [AAB27138](#) Reports Conserved Domains, BLink, Links  
SmN [Mus sp.]  
gi|386203|gb|AAB27138.1|bbm|308258|bbs|133633[386203]

6: [Q17QN3](#) Reports Conserved Domains, BLink, Links

Top Organisms [Tree]

- Homo sapiens (563)
- Mus musculus (363)
- Rattus norvegicus (166)
- Drosophila melanogaster (105)
- Danio rerio (103)
- All other taxa (235)
- More...

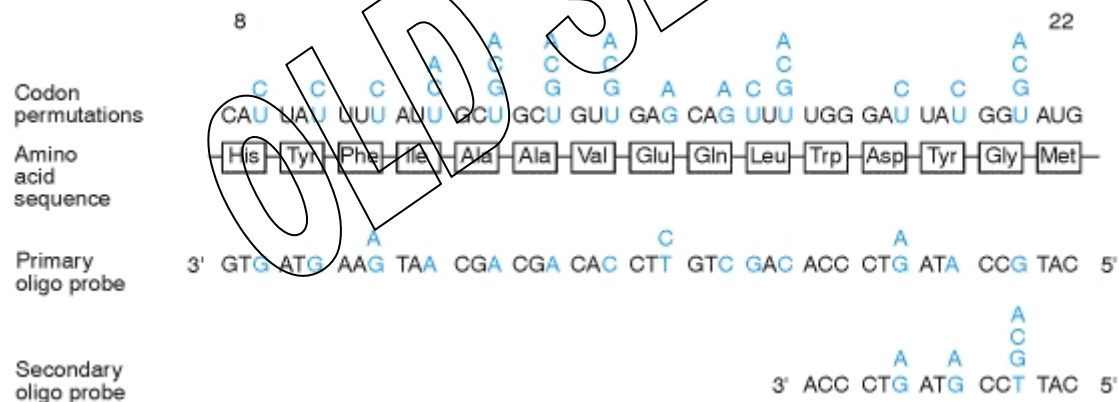
Recent Activity

- (smn sapiens) AND (smn sapiens) (563)
- RecName: Full=Small nuclear ribonucleoprotein-associated protein N; Short=snRNP-N; Short=S...
- smn sapiens (595)
- smn (3605) Protein
- Clinical and genetic analysis of long QT syndrome in children from six families in Saudi A...
- » See more...

# Position-independent gene identification

## [1. Starting from the protein product]

- Classically, the identification of the gene, starting from the protein sequence, was obtained by reverse-translation of the aa sequence, followed by Southern blot



From: Strachan and Read  
Human Molecular Genetics

# Position-independent gene identification

## [1. Starting from the protein product]

- Nowadays, we would BLAST public databases with the sequence of interest...
- ...and the whole experimental strategy would last only a few minutes



# Position-independent gene identification

## [1. Starting from the protein product]

What if you have only the aa sequence?

MASTERSINGENETICSARETHEFINEST

NCBI HomePage - Windows Internet Explorer

http://www.ncbi.nlm.nih.gov/

File Edit View Favorites Tools Help

NCBI HomePage

NCBI

National Center for Biotechnology Information

National Library of Medicine National Institutes of Health

PubMed All Databases BLAST OMIM Books TaxBrowser Structure

Search All Databases for Go

**SITE MAP**  
 Alphabetical List  
 Resource Guide

**About NCBI**  
 An introduction to NCBI

**GenBank**  
 Sequence submission support and software

**Literature databases**  
 PubMed, OMIM, Books, and PubMed Central

**Molecular databases**  
 Sequences, structures, and taxonomy

**Genomic biology**  
 The human genome, whole genomes, and related resources

**Tools**  
 Data mining

**Research at**

**Try NCBI's new home page! NEW**

NCBI's [new home page and site guide](#) provides a catalog of NCBI resources, a consistent menu at the top of each page, and step-by-step instructions for common tasks. The page you are now reading will soon be replaced by the new design.

**What does NCBI do?**

Established in 1988 as a national resource for molecular biology information, NCBI creates public databases, conducts research in computational biology, develops software tools for analyzing genome data, and disseminates biomedical information - all for the better understanding of molecular processes affecting human health and disease. [More about NCBI...](#)

**NLM/NCBI H1N1 Flu Resources**

- [Newest H1N1 influenza sequences](#)
- [Submit flu sequences to GenBank](#)
- [Latest H1N1 citations in PubMed](#)
- [MedlinePlus \(consumer health information\)](#)
- [Enviro-Health links](#)

**Hot Spots**

- Clusters of orthologous groups
- Coffee Break, Genes & Disease, NCBI Handbook
- Electronic PCR
- Entrez Home
- Entrez Tools
- Gene expression omnibus (GEO)
- Human genome resources
- Influenza Virus Resource
- Map Viewer
- dbMHC
- Mouse genome resources
- My NCBI
- ORF finder

BLAST: Basic Local Alignment Search Tool - Windows Internet Explorer

http://blast.ncbi.nlm.nih.gov/Blast.cgi

File Edit View Favorites Tools Help

BLAST: Basic Local Alignment Search Tool

BLAST Basic Local Alignment Search Tool

Home Recent Results Saved Strategies Help

My NCBI Welcome crivolta. [Sign Out]

NCBI/ BLAST Home

BLAST finds regions of similarity between biological sequences. [more...](#)

**New** Aligning Multiple Protein Sequences? Try the **COBALT Multiple Alignment Tool**. [Go](#)

### BLAST Assembled Genomes

Choose a species genome to search, or [list all genomic BLAST databases](#).

- [Human](#)
- [Mouse](#)
- [Rat](#)
- [Arabidopsis thaliana](#)
- [Oryza sativa](#)
- [Bos taurus](#)
- [Danio rerio](#)
- [Drosophila melanogaster](#)
- [Gallus gallus](#)
- [Pan troglodytes](#)
- [Microbes](#)
- [Apis mellifera](#)

### Basic BLAST

Choose a BLAST program to run.

- [nucleotide blast](#) Search a **nucleotide** database using a **nucleotide** query  
*Algorithms:* blastn, megablast, discontinuous megablast
- [protein blast](#) Search **protein** database using a **protein** query  
*Algorithms:* blastp, psi-blast, phi-blast
- [blastx](#) Search **protein** database using a **translated nucleotide** query
- [tblastn](#) Search **translated nucleotide** database using a **protein** query
- [tblastx](#) Search **translated nucleotide** database using a **translated nucleotide** query

### Specialized BLAST

Choose a type of specialized search (or database name in parentheses:)

- Make specific primers with [Primer-BLAST](#)
- Search [trace archives](#)

**News**

[Limit by organism improved](#)

There is a new feature to include or exclude multiple organisms from a search.  
Mon, 14 Sep 2009 09:00:00 EST

[More BLAST news...](#)

**Tip of the Day**

[Use Genomic BLAST to see the genomic context](#)

If you are interested in the evolution of a particular gene or gene family it is often interesting to examine the intro-exon structure even across species.

[More tips...](#)

Protein BLAST: search protein databases using a protein query - Windows Internet Explorer

http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&BLAST\_PROGRAMS=blastp&PAGE\_TYPE=BlastSearch&SHOW\_DEFAULTS=on&LINK\_LOC=blasthome

File Edit View Favorites Tools Help

Protein BLAST: search protein databases using a prot...

**BLAST** Basic Local Alignment Search Tool

Home Recent Results Saved Strategies Help

My NCBI  
Welcome crivolta. [Sign Out]

NCBI/BLAST/blastp suite

blastn blastp blastx tblastn tblastx

Enter Query Sequence

BLASTP programs search protein databases using a protein query. [more...](#)

Reset page Bookmark

Enter accession number, gi, or FASTA sequence

MASTERSINGENETICSARETHEFINEST

Clear

Query subrange

From

To

Or, upload file

Browse...

Job Title

Enter a descriptive title for your BLAST search

☐ Align two or more sequences

Choose Search Set

Database

Non-redundant protein sequences (nr)

Organism

Optional

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown.

☐ Exclude

Entrez Query

Optional

Enter an Entrez query to limit search

Program Selection

Algorithm

☒ blastp (protein-protein BLAST)

☐ PSI-BLAST (Position-Specific Iterated BLAST)

☐ PHI-BLAST (Pattern Hit Initiated BLAST)

Choose a BLAST algorithm

**BLAST**

Search database Non-redundant protein sequences (nr) using Blastp (protein-protein BLAST)

☐ Show results in a new window

Algorithm parameters

General Parameters

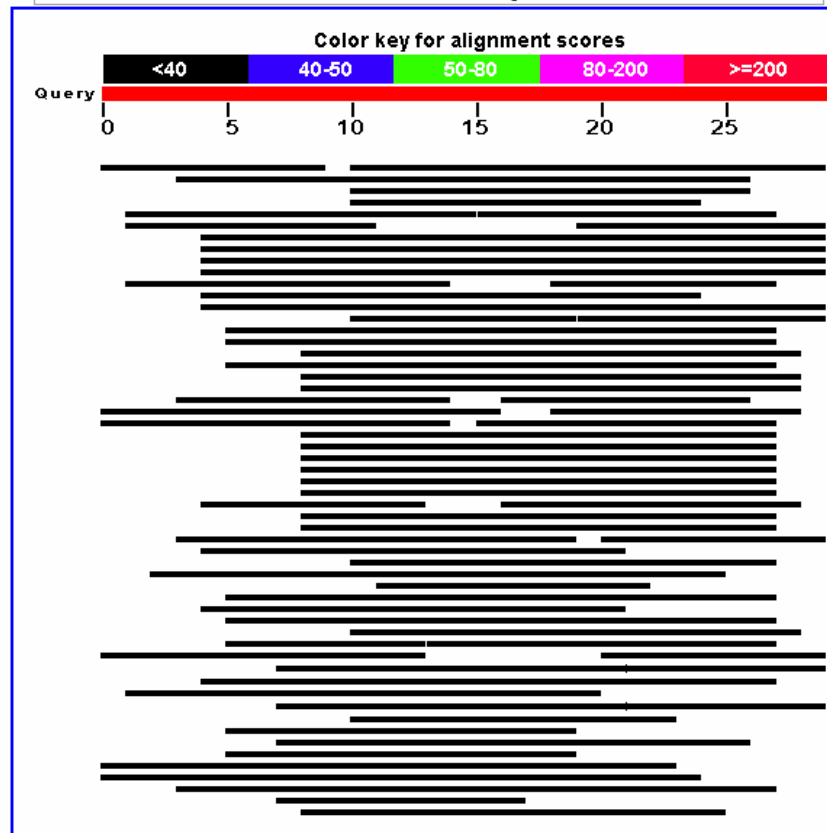
Max target

100

Note: Parameter values that differ from the default are highlighted in yellow

### Distribution of 108 Blast Hits on the Query Sequence

Mouse-over to show define and scores, click to show alignments



[Distance tree of results](#) NEW

Sequences producing significant alignments:

		Score (Bits)	E Value
<a href="#">ref ZP_05706329.1 </a>	transcriptional regulator [Cardiobacterium...]	33.3	3.0
<a href="#">gb AAT28449.1 </a>	envelope glycoprotein [Human immunodeficiency ...]	31.6	9.7
<a href="#">ref ZP_05319723.1 </a>	transcriptional regulator, arsr family [Ne...]	31.2	13

Protein - transcriptional regulator [Cardiobacterium hominis... - Windows Internet Explorer

http://www.ncbi.nlm.nih.gov/sites/entrez?cmd=Retrieve&db=protein&dopt=GenPept&RID=CBPT4UH015&log%24=protop&blast\_rank=1&list\_uids=258546095

File Edit View Favorites Tools Help

Protein - transcriptional regulator [Cardiobacterium ho...

NCBI Protein

All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books

Search Protein for Go Clear

Limits Preview/Index History Clipboard Details

Format: GenPept FASTA Graphics More Formats

Download Save Links

NCBI Reference Sequence: ZP\_05706329.1

**transcriptional regulator [Cardiobacterium hominis ATCC 15826]**

Comment Features Sequence

LOCUS ZP\_05706329 172 aa linear BCT 14-SEP-2009

DEFINITION transcriptional regulator [Cardiobacterium hominis ATCC 15826].

ACCESSION ZP\_05706329

VERSION ZP\_05706329.1 GI:258546095

DBLINK Project: 37269

DBSOURCE REFSEQ: accession NZ\_ACKY01000128.1

KEYWORDS

SOURCE Cardiobacterium hominis ATCC 15826

ORGANISM Cardiobacterium hominis ATCC 15826

Bacteria; Proteobacteria; Gammaproteobacteria; Cardiobacteriales; Cardiobacteriaceae; Cardiobacterium.

REFERENCE 1 (residues 1 to 172)

AUTHORS Qin,X., Bachman,B., Battles,P., Bell,A., Bess,C., Bickham,C., Chaboub,L., Chen,D., Coyle,M., Deiros,D.R., Dinh,H., Forbes,L., Fowler,G., Francisco,L., Fu,Q., Gubbala,S., Hale,W., Han,Y., Hemphill,L., Highlander,S.K., Hirani,K., Hognes,M., Jackson,L., Jakkamsetti,A., Javaid,M., Jiang,H., Korchina,V., Kovar,C., Lara,F., Lee,S., Mata,R., Mathew,T., Moen,C., Morales,K., Munidasa,M., Nazareth,L., Ngo,R., Nguyen,L., Okwuonu,G., Ongerli,F., Patil,S., Petrosino,J., Pham,C., Pham,P., Pu,L.-L., Puazo,M., Raj,R., Reid,J., Rouhana,J., Saada,N., Shang,Y., Simmons,D., Thornton,R., Warren,J., Weissenberger,G., Zhang,J., Zhang,L., Zhou,C., Zhu,D., Muzny,D., Worley,K. and Gibbs,R.

TITLE Direct Submission

JOURNAL Submitted (10-AUG-2009) Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 2 (residues 1 to 172)

AUTHORS Qin,X., Bachman,B., Battles,P., Bell,A., Bess,C., Bickham,C., Chaboub,L., Chen,D., Coyle,M., Deiros,D.R., Dinh,H., Forbes,L., Fowler,G., Francisco,L., Fu,Q., Gubbala,S., Hale,W., Han,Y., Hemphill,L., Highlander,S.K., Hirani,K., Hognes,M., Jackson,L., Jakkamsetti,A., Javaid,M., Jiang,H., Korchina,V., Kovar,C., Lara,F., Lee,S., Mata,R., Mathew,T., Moen,C., Morales,K., Munidasa,M., Nazareth,L., Ngo,R., Nguyen,L., Okwuonu,G., Ongerli,F., Patil,S., Petrosino,J., Pham,C., Pham,P., Pu,L.-L., Puazo,M.,

Change Region Shown

Customize View

**BLAST Sequence**

Find regions of similarity between this sequence and other sequences using BLAST.

**Conserved Domains**

View conserved domains detected in this protein sequence using CD-search.

**Identical Proteins for ZP\_05706329.1**

transcriptional regulator [Cardiobacterium h... [EEV87553]

» See all...

**Recent Activity**

Turn Off Clear

transcriptional regulator [Cardiobacterium hominis ATCC 15826]

Protein Sequence (29 lett...

Protein Sequence (29 lett... BLAST

(smn sapiens) AND "Homo s... (563) Protein

RecName: Full=Small nuclear ribonucleoprotein-associated protein N; Short=snRNP-N;

» See more...

**All links from this record**

Conserved Domains

# Position-independent gene identification

## [2. Starting from an animal model]

- Another (relatively rare) possibility comes from the identification of the gene in a mouse model and subsequent identification in another species (e.g. human)
- Again, in the past this was done with DNA probes. Today, we would again use BLAST.

# Position-independent gene identification

## [3. The “candidate gene” approach]

- The choice of candidates is based on:
  1. Tissue specificity and abundance
  2. Chromosomal location
  3. Sequence information and length
  4. Similarity to other known disease-causing genes
  5. Other characteristics (gene conservation, etc.)



Database - Windows Internet Explorer

http://www.ncbi.nlm.nih.gov/Database/index.html

File Edit View Favorites Tools Help

Database

NCBI **Databases**

PubMed Entrez BLAST OMIM Books TaxBrowser Structure

Search Entrez for  Go

NCBI

Site Map PubMed  
Guide to Protein  
resource Nucleotide  
EST  
GSS  
Structure  
Entrez+ Genome  
Help docu Genome  
for the En Books  
Entrez T CancerChromosomes  
Conserved Domains  
3D Domains  
Gene  
Entrez G Genome Project  
Query dbGaP  
Search a GENSAT  
Entrez da GEO Profiles  
GEO Datasets  
Entrez T HomoloGene  
Links to a Journals  
Entrez too MeSH  
Batch Ent NLM Catalog  
Utilities OMIA  
OMIM  
NCBI Ha PMC  
In-depth g PopSet  
NCBI reso Probe  
Protein Clusters

The integrated, text-based search and retrieval system used for the major databases, including PubMed, Nucleotide and sequences, Protein Structures, Complete Genomes, and others. Click on the graphic below for a more view of Entrez integration.

The complete list of Entrez databases can be viewed in the search pull down menu.

**Nucleotide Databases**

[dbEST](#) [PopSet](#)  
[dbGSS](#) [Probe](#)

smn sapiens - Gene Results - Windows Internet Explorer

http://www.ncbi.nlm.nih.gov/sites/entrez

File Edit View Favorites Tools Help

smn sapiens - Gene Results

NCBI Entrez Gene

My NCBI  
Welcome crivolta. [Sign Out]

All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books

Search Gene for smn sapiens Go Clear Save Search

Limits Preview/Index History Clipboard Details

Display Summary Show 20 Sort by Relevance Send to

All: 70 Current Only: 64 Genes Genomes: 64 SNP GeneView: 60

Items 1 - 20 of 70 Page 1 of 4 Next

☐ 1: [SMN1](#) Order cDNA clone, Links  
**Official Symbol** SMN1 and **Name:** survival of motor neuron 1, telomeric [*Homo sapiens*]  
**Other Aliases:** BCD541, SMA, SMA1, SMA2, SMA3, SMA4, SMA@, SMN, SMNT, T-BCD541  
**Other Designations:** gemin 1  
**Chromosome:** 5; **Location:** 5q13  
**Annotation:** Chromosome 5, NC\_000005.9 (70220768..70248839)  
**MIM:** 600444  
**GeneID:** 16

☐ 2: [SNRPN](#) Order cDNA clone, Links  
**Official Symbol** SNRPN and **Name:** small nuclear ribonucleoprotein polypeptide N [*Homo sapiens*]  
**Other Aliases:** DKFZp686C0927, DKFZp686M12165, DKFZp76111912, DKFZp762N022, FLJ33569, FLJ36996, FLJ39265, HCERN3, MGC29886, PWCR, RT-LI, SM-D, SMN, SNRNP-N, SNURF-SNRPN  
**Other Designations:** SM protein N; tissue-specific splicing protein  
**Chromosome:** 15; **Location:** 15q11.2  
**Annotation:** Chromosome 15, NC\_000015.9 (25068794..25664609)  
**MIM:** 182279  
**GeneID:** 6638

☐ 3: [STMN1](#) Order cDNA clone, Links  
**Official Symbol** STMN1 and **Name:** stathmin 1 [*Homo sapiens*]  
**Other Aliases:** RP1-125I3.5, C1orf215, FLJ32206, LAP18, Lag, MGC138869, MGC138870, OP18, PP17, PP19, PR22, SMN  
**Other Designations:** OTTHUMP00000008530; leukemia-associated phosphoprotein p18; metastastin; oncoprotein 18; phosphoprotein 19; prosolin; stathmin 1/oncoprotein 18  
**Chromosome:** 1; **Location:** 1p36.1-p35

**Recent Activity**  
 Turn Off Clear  
 Q [smn sapiens](#) (70)  
 Q [smn](#) (1297) Gene  
 Q [Nucleotide for Protein \(S...](#) (2) Nucleotide  
 Q [transcriptional regulator \[Cardiobacterium hominis ATCC 15826\]](#)  
 Q [Protein Sequence \(29 lett...](#) BLAST  
 » See more...

Entrez Gene: SMN1 survival of motor neuron 1, telomeric [ Homo sapiens ] - Windows Internet Explorer

http://www.ncbi.nlm.nih.gov/gene/6606?ordinalpos=1&itool=EntrezSystem2.PEntrez.Gene.Gene\_ResultsPanel.Gene\_RVDocSum

File Edit View Favorites Tools Help

Entrez Gene: SMN1 survival of motor neuron 1, telom...

NCBI Entrez Gene

All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books

Search Gene for Go Clear

Limits Preview/Index History Clipboard Details

Display Full Report Show 20 Sort by Relevance Send to

All: 1 Current Only: 1 Genes Genomes: 1 SNP GeneView: 1

1: SMN1 survival of motor neuron 1, telomeric [ Homo sapiens ]  
GeneID: 6606 updated 06-Sep-2009

Summary

**Official Symbol** SMN1 provided by HGNC

**Official Full Name** survival of motor neuron 1, telomeric provided by HGNC

**Primary source** HGNC:11117

**See related** Ensembl:ENSG00000172062; HPRD:02646; MIM:600354

**Gene type** protein coding

**RefSeq status** REVIEWED

**Organism** Homo sapiens

**Lineage** Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo

**Also known as** SMA; SMN; SMA1; SMA2; SMA3; SMA4; SMA@; SMNT; BCD541; T-BCD541; SMN1

**Summary** This gene is part of a 500 kb inverted duplication on chromosome 5q13. This duplicated region contains at least four genes and repetitive elements which make it prone to rearrangements and deletions. The repetitiveness and complexity of the sequence have also caused difficulty in determining the organization of this genomic region. The telomeric and centromeric copies of this gene are nearly identical and encode the same protein. However, mutations in this gene, the telomeric copy, are associated with spinal muscular atrophy; mutations in the centromeric copy do not lead to disease. The centromeric copy may be a modifier of disease caused by mutation in the telomeric copy. The critical sequence difference between the two genes is a single nucleotide in exon 7, which is thought to be an exon splice enhancer. Note that the nine exons of both the telomeric and centromeric copies are designated historically as exon 1, 2a, 2b, and 3-8. It is thought that gene conversion events may involve the two genes, leading to varying copy numbers of each gene. The protein encoded by this gene localizes to both the cytoplasm and the nucleus. Within the nucleus, the protein localizes to subnuclear bodies called gems which are found near coiled bodies containing high concentrations of small ribonucleoproteins (snRNPs). This protein forms heteromeric complexes with proteins such as SIP1 and GEMIN4, and also interacts with several proteins known to be involved in the biogenesis of snRNPs, such as hnRNP U protein and the small nucleolar RNA binding protein. Two transcript variants encoding distinct isoforms have been described. [provided by RefSeq]

Entrez Gene Home

Table Of Contents

- Summary
- Genomic regions, transcripts...
- Genomic context
- Bibliography
- Interactions
- General gene information
- General protein information
- Reference Sequences
- Related Sequences
- Additional Links

Links

- Order cDNA clone
- BioSystems
- Books
- CCDS
- Conserved Domains
- Genome
- GEO Profiles
- HomoloGene
- Map Viewer
- Nucleotide
- OMIM
- BioAssay, by Gene target
- PubChem Compound
- PubChem Substance
- Full text in PMC
- Probe
- Protein
- PubMed
- PubMed (OMIM)
- PubMed (GeneRIF)
- SNP
- SNP: Genotype
- SNP: GeneView

Done

start splicing-biochemi... Eudora - [In] 2009-2010 Carlo\_Rivolta\_Ma... Carlo\_Rivolta\_Ma... Entrez Gene: SM... EditPad Lite

Entrez Gene: SMN1 survival of motor neuron 1, telomeric [ Homo sapiens ] - Windows Internet Explorer

http://www.ncbi.nlm.nih.gov/gene/6606?ordinalpos=1&tool=EntrezSystem2.PEntrez.Gene.Gene\_ResultsPanel.Gene\_RVDocSum

File Edit View Favorites Tools Help

Entrez Gene: SMN1 survival of motor neuron 1, telom...

Primates; Haplorrhini; Catarrhini; Hominidae; Homo

**Also known as** SMA; SMN; SMA1; SMA2; SMA3; SMA4; SMA@; SMNT; BCD541; T-BCD541; SMN1

**Summary** This gene is part of a 500 kb inverted duplication on chromosome 5q13. This duplicated region contains at least four genes and repetitive elements which make it prone to rearrangements and deletions. The repetitiveness and complexity of the sequence have also caused difficulty in determining the organization of this genomic region. The telomeric and centromeric copies of this gene are nearly identical and encode the same protein. However, mutations in this gene, the telomeric copy, are associated with spinal muscular atrophy; mutations in the centromeric copy do not lead to disease. The centromeric copy may be a modifier of disease caused by mutation in the telomeric copy. The critical sequence difference between the two genes is a single nucleotide in exon 7, which is thought to be an exon splice enhancer. Note that the nine exons of both the telomeric and centromeric copies are designated historically as exon 1, 2a, 2b, and 3-8. It is thought that gene conversion events may involve the two genes, leading to varying copy numbers of each gene. The protein encoded by this gene localizes to both the cytoplasm and the nucleus. Within the nucleus, the protein localizes to subnuclear bodies called gems which are found near coiled bodies containing high concentrations of small ribonucleoproteins (snRNPs). This protein forms heteromeric complexes with proteins such as SIP1 and GEMIN4, and also interacts with several proteins known to be involved in the biogenesis of snRNPs, such as hnRNP U protein and the small nucleolar RNA binding protein. Two transcript variants encoding distinct isoforms have been described. [provided by RefSeq]

**Genomic regions, transcripts, and products**

Go to [reference sequence details](#) [Try our new Sequence Viewer](#)

**NC\_000005.9**

[ 70220768 5' ] [ 70248839 3' ]

NM\_000344.3  
NM\_022874.2

NP\_000335.1 isoform d  
NP\_075012.1 isoform b

CCDS34181.1  
CCDS34182.1

■ - coding region ■ - untranslated region

**Genomic context**

chromosome: 5; Location: 5q13 [See SMN1 in MapViewer](#)

[ 70156009 ] [ 70320941 ]

LOC100933625 SERP1A SMN1 LOC100287378 NRIP

**Bibliography**

Related Articles in PubMed

**Books**  
CCDS  
Conserved Domains  
Genome  
GEO Profiles  
HomoloGene  
Map Viewer  
Nucleotide  
OMIM  
BioAssay, by Gene target  
PubChem Compound  
PubChem Substance  
Full text in PMC  
Probe  
Protein  
PubMed  
PubMed (OMIM)  
PubMed (GeneRIF)  
SNP  
SNP: Genotype  
SNP: GeneView  
Taxonomy  
UniSTS  
AceView  
Ensembl  
Evidence Viewer  
GeneTests for MIM: 253300  
GeneTests for MIM: 600354  
HGNC  
HPRD  
HuGE Navigator  
KEGG  
MGC  
ModelMaker  
SMN1  
UCSC  
BioAssay  
UniGene  
LinkOut

**Entrez Gene Info**

**Feedback**

**Subscriptions**

Survival of motor neuron 1, telomeric (SMN1) - Windows Internet Explorer

http://www.ncbi.nlm.nih.gov/UniGene/clust.cgi?UGID=1372102&TAXID=9606&SEARCH=

File Edit View Favorites Tools Help

Survival of motor neuron 1, telomeric (SMN1)

NCBI

UniGene  
ORGANIZED VIEW OF THE TRANSCRIPTOME

PubMed Nucleotide **Protein** Genome Structure Popset Taxonomy

Search UniGene Go Clear

UGID:1372102 UniGene Hs.535788 Homo sapiens (human) SMN1 Order cDNA clone, Links

### Survival of motor neuron 1, telomeric (SMN1)

#### SELECTED PROTEIN SIMILARITIES

Comparison of sequences in UniGene with selected protein reference sequences. The alignments can suggest function of a gene.

	Reference Protein	Species	Id(%)	Len(aa)
<a href="#">XP_001156488.1</a>	PREDICTED: survival of motor neuron 1, telomeric isoform 7	<i>P. troglodytes</i>	100.0	296
<a href="#">NP_059107.1</a>	survival of motor neuron 2, centromeric isoform d	<i>H. sapiens</i>	100.0	293

#### GENE EXPRESSION

Tissues and development stages from this gene's sequences survey gene expression. Links to other NCBI expression resources.

**EST Profile:** Approximate expression patterns inferred from EST sources.  
[Show more entries with profiles like this]

**GEO profile:** Experimental gene expression data (Gene Expression Omnibus).

**cDNA Sources:** stomach; uncharacterized tissue; intestine; brain; mixed; mouth; blood; lung; prostate; placenta; ovary; mammary gland; pancreas; uterus; adrenal gland; skin; amniotic fluid; eye; cervix; lymph; heart; connective tissue; kidney; embryonic tissue; lymph node; thyroid; umbilical cord; testis; liver

#### MAPPING POSITION

Genomic location identified by transcript mapping, radiation hybrid mapping, genetic mapping or cytogenetic mapping.

Chromosome:	5
Map position:	5q13
UniSTS entry:	Chr 5 <a href="#">G20836</a>
UniSTS entry:	Chr 5 <a href="#">A006111</a>
UniSTS entry:	Chr 5 <a href="#">D5S1946</a>
UniSTS entry:	Chr 5 <a href="#">D5S2463</a>
UniSTS entry:	Chr 5 <a href="#">STS-U18423</a>
UniSTS entry:	Chr 5 <a href="#">D5S2798</a>

start splicing-biochemistry Eudora - [In] 2009-2010 Carlo\_Rivolta\_Master... Survival of motor neu... EditPad Lite Internet 100% 11:53

EST Profile - Hs.535788 - Windows Internet Explorer

http://www.ncbi.nlm.nih.gov/UniGene/ESTProfileViewer.cgi?uglist=Hs.535788

File Edit View Favorites Tools Help

EST Profile - Hs.535788

NCBI UniGene EST Profile Viewer

Pubmed Nucleotide Protein Genome Structure Popset Taxonomy

Search UniGene

### EST Profile

Hs.535788 - SMN1: Survival of motor neuron 1, telomeric

#### Breakdown by Body Sites

		Hs.535788	
adipose tissue	0	0 / 13159	
adrenal gland	30	1 / 33324	
ascites	0	0 / 40067	
bladder	0	0 / 30121	
blood	24	3 / 123959	
bone	0	0 / 71802	
bone marrow	0	0 / 48949	
brain	12	14 / 1104257	
cervix	61	3 / 48506	
connective tissue	40	6 / 149531	
ear	0	0 / 16345	
embryonic tissue	4	1 / 215831	
esophagus	0	0 / 20212	
eye	18	4 / 211771	
heart	44	4 / 90351	
intestine	21	5 / 235458	
kidney	28	6 / 212588	
larynx	0	0 / 24438	
liver	4	1 / 208370	
lung	35	12 / 338117	
lymph	22	1 / 44401	
lymph node	10	1 / 91923	
mammary gland	51	8 / 154363	
mouth	178	12 / 67225	
muscle	0	0 / 108174	
nerve	0	0 / 15820	
ovary	19	2 / 102623	

Survival of motor neuron 1, telomeric (SMN1) - Windows Internet Explorer

http://www.ncbi.nlm.nih.gov/UniGene/clust.cgi?UGID=1372102&TAXID=9606&SEARCH=

File Edit View Favorites Tools Help

Survival of motor neuron 1, telomeric (SMN1)

NCBI

UniGene  
ORGANIZED VIEW OF THE TRANSCRIPTOME

PubMed Nucleotide **Protein** Genome Structure Popset Taxonomy

Search UniGene Go Clear

UGID:1372102 UniGene Hs.535788 Homo sapiens (human) SMN1 Order cDNA clone, Links

### Survival of motor neuron 1, telomeric (SMN1)

#### SELECTED PROTEIN SIMILARITIES

Comparison of sequences in UniGene with selected protein reference sequences. The alignments can suggest function of a gene.

Reference Protein	Species	Id(%)	Len(aa)
<a href="#">XP_001156488.1</a> PREDICTED: survival of motor neuron 1, telomeric isoform 7	<i>P. troglodytes</i>	100.0	296
<a href="#">NP_059107.1</a> survival of motor neuron 2, centromeric isoform d	<i>H. sapiens</i>	100.0	293

#### GENE EXPRESSION

Tissues and development stages from this gene's sequences survey gene expression. Links to other NCBI expression resources.

**EST Profile:** Approximate expression patterns inferred from EST sources.  
[Show more entries with profiles like this]

**GEO profiles:** Experimental gene expression data (Gene Expression Omnibus).

**cDNA Sources:** stomach; uncharacterized tissue; intestine; brain; mixed; mouth; blood; lung; prostate; placenta; ovary; mammary gland; pancreas; uterus; adrenal gland; skin; amniotic fluid; eye; cervix; lymph; heart; connective tissue; kidney; embryonic tissue; lymph node; thyroid; umbilical cord; testis; liver

#### MAPPING INFORMATION

Genomic locations specified by transcript mapping, radiation hybrid mapping, genetic mapping or cytogenetic mapping.

Chromosome	Map position	UniSTS entry
5	5q13	
	Chr 5	<a href="#">G20836</a>
	Chr 5	<a href="#">A006111</a>
	Chr 5	<a href="#">D5S1946</a>
	Chr 5	<a href="#">D5S2463</a>
	Chr 5	<a href="#">STS-U18423</a>
	Chr 5	<a href="#">D5S2798</a>

start

splicing-biochemistry

Eudora - [In]

2009-2010

Carlo\_Rivolta\_Master...

Survival of motor neu...

EditPad Lite

Internet

100%

11:53



GEO Profile Links for UniGene (UniGene Links for Gene (Select 6606)) - GEO Profiles Results - Windows Internet Explorer

http://www.ncbi.nlm.nih.gov/sites/entrez?DB=geo&DbFrom=unigene&FromResult=1372102&cmd=Link&LinkName=unigene\_geo&tool=UniGene.clust

File Edit View Favorites Tools Help

GEO Profile Links for UniGene (UniGene Links for Gene...)

My NCBI  
Welcome crivolta. [Sign Out]

All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books

Search GEO Profiles for [ ] Go Clear

Limits Preview/Index History Clipboard Details

Display Summary Show 20 Sort By Send to Download profile data

All: 17

Items 1 - 17 of 17 One page.

☐ 1: GDS1839 record | GPL3113 20796 [ *Homo sapiens* ] 24 samples Profile Neighbors, Chromosome Neighbors, Homologs, Links

Annotation: SMN1: Survival of motor neuron 1, telomeric (multiple annotations exist)

Reporter: R81675 R81676 IMAGE:147738 (SPOT ID)

Experiment: Follicular lymphomas and response to rituximab, gene expression array-based, log2 ratio

☐ 2: GDS1696 record | GPL2895 375008 [ *Homo sapiens* ] 12 samples Profile Neighbors, Chromosome Neighbors, Sequence Neighbors, Homologs, Links

Annotation: SMN2: survival of motor neuron 2, centromeric BCD541, C-BCD541, FLJ76644, MGC20996, MGC5208, SMNC (multiple annotations exist)

Reporter: NM\_022874.1 NM\_022877.1 NM\_022876.1 NM\_000344.2 NM\_017411.2 NM\_022875.1 BX399610.2

Experiment: Nanosecond pulsed electric fields effect: time course, gene expression array-based, count

☐ 3: GDS85 record | GPL169 9501 [ *Homo sapiens* ] 10 samples Profile Neighbors, Chromosome Neighbors, Homologs, Links

Annotation: SMN1: Survival of motor neuron 1, telomeric

Reporter: AA029189 IMAGE:470039 (clone)

Experiment: Serum stimulation time course (10k\_print2), gene expression array-based, log ratio

☐ 4: GDS1761 record | GPL1290 7819 [ *Homo sapiens* ] 64 samples Chromosome Neighbors, Homologs, Links

Annotation: SMN1: Survival of motor neuron 1, telomeric

Reporter: AA029189 AA029190 IMAGE:470039 (clone)

NCBI  
Home  
FAQ  
Browse  
Downloads  
Submit  
Graph caption  
Contact Us

Entrez  
Global Search  
About  
Help



SMN1 Gene - GeneCards | SMN Protein | SMN Antibody - Windows Internet Explorer

http://www.genecards.org/cgi-bin/carddisp.pl?gene=SMN1&search=smn1

File Edit View Favorites Tools Help

SMN1 Gene - GeneCards | SMN Protein | SMN Antibody

(Experimental results according to [GeneNote](#) and [GNF SymAtlas](#), probe sets-to-genes annotations according to [GeneAnnot](#), [GeneTide](#), Sets of similar genes according to [GeneDecks](#), Electronic Northern calculations according to data from [UniGene](#) (Build 219 Homo sapiens), [SAGE](#) tags according to [CGAP](#), plus additional links to [SOURCE](#), and/or [GNF BioGPS](#), and/or [EXPLODB](#), and/or [UniProtKB](#), [Expression Assays from Applied Biosystems](#) )

[About This Section](#)

Jump to Section...

Expression according to [GeneNote](#) / [GeneAnnot](#) / [GeneTide](#)

3 probe-sets matching SMN1 gene

Affymetrix probe-set	Array	GeneAnnot data			GeneNote data		GeneTide data				
		# genes	Sensitivity	Specificity	Correlation	Length	Gb Accession	Consensus	Uniqueness	Score	Rank
37313_at <sup>2,3</sup>	U95-A	2	1.00	0.50	1.00	1.00	U80017	0.20	0.14	0.17	1
203852_s_at <sup>2,3</sup>	U133-A	2	1.00	0.50	--	--	NM_000344	0.40	0.44	0.42	1
203852_s_at <sup>2</sup>	U133Plus2	2	1.00	0.50	--	--	--	--	--	--	--

[About this table](#)

Data from [GeneNote](#) ([Publications](#)) and [GNF SymAtlas](#) (See also [BioGPS](#))

[About these images](#)

www.genecards.org

Survival of motor neuron 1, telomeric (SMN1) - Windows Internet Explorer

http://www.ncbi.nlm.nih.gov/UniGene/clust.cgi?UGID=1372102&TAXID=9606&SEARCH=

File Edit View Favorites Tools Help

Survival of motor neuron 1, telomeric (SMN1)

NCBI

UniGene  
ORGANIZED VIEW OF THE TRANSCRIPTOME

PubMed Nucleotide Protein Genome Structure Popset Taxonomy

Search UniGene Go Clear

UGID:1372102 UniGene Hs.535788 Homo sapiens (human) SMN1 Order cDNA clone, Links

### Survival of motor neuron 1, telomeric (SMN1)

#### SELECTED PROTEIN SIMILARITIES

Comparison of sequences in UniGene with selected protein reference sequences. The alignments can suggest function of a gene.

Reference Protein	Species	Id(%)	Len(aa)
<a href="#">XP_001156488.1</a> PREDICTED: survival of motor neuron 1, telomeric isoform 7	<i>P. troglodytes</i>	100.0	296
<a href="#">NP_059107.1</a> survival of motor neuron 2, centromeric isoform d	<i>H. sapiens</i>	100.0	293

#### GENE EXPRESSION

Tissues and development stages from this gene's sequences survey gene expression. Links to other NCBI expression resources.

**EST Profile:** Approximate expression patterns inferred from EST sources.  
[\[Show more entries with profiles like this\]](#)

**GEO profiles:** Experimental gene expression data (Gene Expression Omnibus).

**cDNA Sources:** stomach; uncharacterized tissue; intestine; brain; mixed; mouth; blood; lung; prostate; placenta; ovary; mammary gland; pancreas; uterus; adrenal gland; skin; amniotic fluid; eye; cervix; lymph; heart; connective tissue; kidney; embryonic tissue; lymph node; thyroid; umbilical cord; testis; liver

#### MAPPING POSITION

Genomic location specified by transcript mapping, radiation hybrid mapping, genetic mapping or cytogenetic mapping.

Chromosome:	5
Map position:	5q13
UniSTS entry:	Chr 5 <a href="#">G20836</a>
UniSTS entry:	Chr 5 <a href="#">A006111</a>
UniSTS entry:	Chr 5 <a href="#">D5S1946</a>
UniSTS entry:	Chr 5 <a href="#">D5S2463</a>
UniSTS entry:	Chr 5 <a href="#">STS-U18423</a>
UniSTS entry:	Chr 5 <a href="#">D5S2798</a>

start

splicing-biochemistry

Eudora - [In]

2009-2010

Carlo\_Rivolta\_Master...

Survival of motor neu...

EditPad Lite

Internet

100%

11:53

Digital Differential Display - Windows Internet Explorer

http://www.ncbi.nlm.nih.gov/UniGene/ddd.cgi?TAX=9606&log\$=BlueSideBar

File Edit View Favorites Tools Help

Digital Differential Display

NCBI

UniGene

ORGANIZED VIEW OF THE TRANSCRIPTOME

PubMed Nucleotide Protein Genome Structure Popset Taxonomy

Search UniGene Go Clear

Limits Preview/Index History Clipboard Details

NCBI

UniGene

Homepage

FAQs

Query Tips

Library Browser

DDD

Download UniGene

Related Databases

Gene

HomoloGene

dbEST

Trace Archive

NIH cDNA Projects

MGC ZGC XGC

Digital Differential Display (DDD)

DDD is a tool for comparing EST profiles in order to identify genes with significantly different expression levels ([More about DDD](#)).

Species: Homo sapiens

Select the taxonomic species of interest from the list above.

Digital Differential Display - Windows Internet Explorer

http://www.ncbi.nlm.nih.gov/UniGene/ddd.cgi?ACT=ok%3A2&TAXID=9606&Pool1=Adipose+tissue%3A10983&Pool2=control%3A7037+7038+8975+11923&DESC=Blood&LID=

File Edit View Favorites Tools Help

Digital Differential Display

FAQs  
Query Tips  
Library Browser  
DDD  
Download UniGene

Related Databases  
Gene  
HomoloGene  
dbEST  
Trace Archive

NIH cDNA Projects  
MGC | ZGC | XGC  
Finding cDNAs

Species: *Homo sapiens* (human) [Start Over](#)

Pool A: Adipose tissue 1 libraries, 8299 ESTs [Edit Pool](#)

Pool B: Blood 4 libraries, 14038 ESTs [Edit Pool](#)

[New Pool](#)

## DDD

### Differential Display Results

The following genes (UniGene entries) display statistically significant differences in EST counts by the Fisher Exact Test.

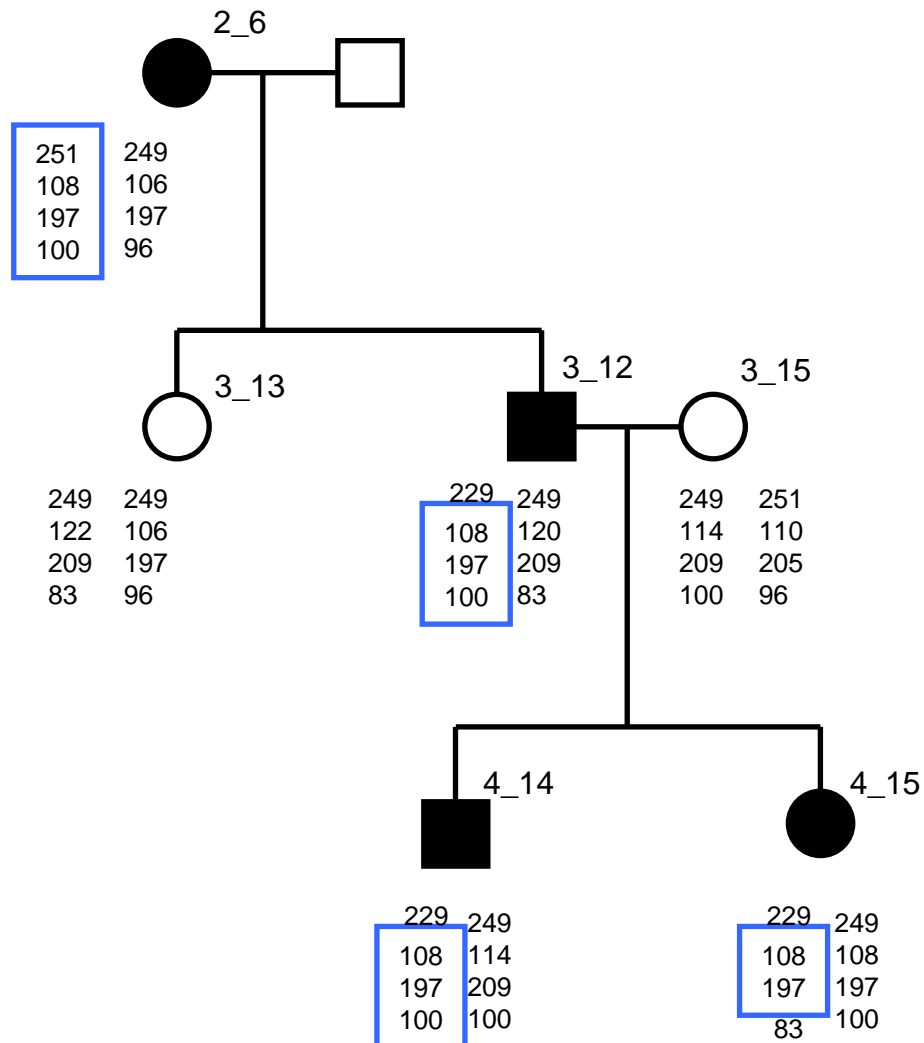
A Adipos..	B Blood	UniGene Entry
0.0000	0.0824	<a href="#">Hs.23118</a> Carbonic anhydrase I (CA1)
0.0012	0.0323	<a href="#">Hs.433670</a> Ferritin, light polypeptide (FTL)
0.0040	0.0338	<a href="#">Hs.709313</a> Beta-2-microglobulin (B2M)
0.0005	0.0209	<a href="#">Hs.523443</a> Hemoglobin, beta (HBB)
0.0025	0.0210	<a href="#">Hs.642877</a> Metastasis associated lung adenocarcinoma transcript 1 (non-protein coding) (MALAT1)
0.0000	0.0163	<a href="#">Hs.449585</a> Immunoglobulin lambda locus (IGL@)
0.0001	0.0155	<a href="#">Hs.432121</a> Peroxiredoxin 2 (PRDX2)
0.0000	0.0123	<a href="#">Hs.46423</a> Histone cluster 1, H4c (HIST1H4C)
0.0004	0.0096	<a href="#">Hs.131226</a> BCL2/adenovirus E1B 19kDa interacting protein 3-like (BNIP3L)
0.0088	0.0002	<a href="#">Hs.517168</a> Transgelin 2 (TAGLN2)
0.0113	0.0031	<a href="#">Hs.586423</a> Eukaryotic translation elongation factor 1 alpha 1 (EEF1A1)
0.0000	0.0073	<a href="#">Hs.155097</a> Carbonic anhydrase II (CA2)

# Position-independent gene identification

## [3. The “candidate gene” approach]

- The choice of candidates is based on:
  1. Tissue specificity and abundance
  2. Chromosomal location
  3. Sequence information and length
  4. Similarity to other known disease-causing genes
  5. Other characteristics (gene conservation, etc.)

# Microsatellite mapping



D10S1573  
D10S595  
D10S1568  
D10S1749

Chr. 10

NCBI HomePage - Windows Internet Explorer

http://www.ncbi.nlm.nih.gov/

File Edit View Favorites Tools Help

Digital Differential Display NCBI HomePage

**NCBI**

**National Center for Biotechnology Information**  
[National Library of Medicine](#) [National Institutes of Health](#)

PubMed All Databases BLAST OMIM Books TaxBrowser Structure

Search  All Databases

**SITE MAP**  
 Alphabetical List  
 Resource Guide

**About NCBI**  
 An introduction to NCBI

**GenBank**  
 Sequence submission support and software

**Literature databases**  
 PubMed, OMIM, Books, and PubMed Central

**Molecular databases**  
 Sequences, structures, and taxonomy

**Genomic biology**  
 The human genome, whole genomes,

**Try NCBI's new home page! NEW**

NCBI's [new home page and site guide](#) provides a catalog of NCBI resources, a consistent menu bar at the top of each page, and step-by-step instructions for common tasks. The page you are now reading will soon be replaced by the new design.

**What does NCBI do?**

Established in 1988 as a national resource for molecular biology information, NCBI creates public databases, conducts research in computational biology, develops software tools for analyzing genome data, and disseminates biomedical information - all for the better understanding of molecular processes affecting human health and disease. [More about NCBI...](#)

**NLM/NCBI H1N1 Flu Resources**

- [Newest H1N1 influenza sequences](#)

**Hot Spots**

- Clusters of orthologous groups
- Coffee Break, Genes & Disease, NCBI Handbook
- Electronic PCR
- Entrez Home
- Entrez Tools
- Gene expression omnibus (GEO)
- Human genome resources
- Influenza Virus Resource
- Map Viewer
- dbMHC

Map Viewer - Windows Internet Explorer

http://www.ncbi.nlm.nih.gov/mapview/

File Edit View Favorites Tools Help

Digital Differential Display Map Viewer

NCBI Home GenBank BLAST

Map Viewer Home Help

The Map Viewer provides a wide variety of genome mapping and sequencing data. [More..](#)

**Search**

Search:  for:   
Go

**Tools Legend**

- Q Search or Browse the Genome
- B BLAST
- Cf Clone Finder
- R Go to region on a chromosome
- G Genome Resources page

**News**

**Human build 37 released** Aug 4  
An update to the human genome assembly and annotation is now... [more](#)  
[Show all](#)

**Related Resources**

- NCBI Home
- NCBI Web Search
- NCBI Site map
- Genome Browser agreement
- Genome Biology
- Taxonomy
- Entrez (Global Query)
- BLAST
- Map Viewer FTP

**Small Genomes**

- Bacteria B
- Organelles B
- Viruses B

**Vertebrates (17)**

**Mammals (14)**

**Primates (3)**

Scientific name	Common name	Build	Tools
<i>Homo sapiens</i>	human	<a href="#">Build 37.1</a>	Q B R G
		<a href="#">Build 36</a>	Q B R Cf
<i>Macaca mulatta</i>	rhesus macaque	<a href="#">Build 1.1</a>	Q B R G
<i>Pan troglodytes</i>	chimpanzee	<a href="#">Build 2.1</a>	Q B R G

**Rodents (2)**

Scientific name	Common name	Build	Tools
<i>Mus musculus</i>	laboratory mouse	<a href="#">Build 37.1</a>	Q B R Cf G
		<a href="#">Build 36.1</a>	Q B R
<i>Rattus norvegicus</i>	rat	<a href="#">RGSC v3.4</a>	Q B R G

**Monotremes (1)**

**Marsupials (1)**

**Other Mammals (7)**

**Other Vertebrates (3)**

**Invertebrates (12)**

**Protozoa (18)**

**Plants (46)**

**Fungi (17)**

Scientific name	Common name	Build	Tools
<i>Aspergillus clavatus</i>		<a href="#">Build 1.1</a>	Q B G
<i>Aspergillus fumigatus</i>		<a href="#">Build 2.1</a>	Q B R G
<i>Aspergillus niger</i>		<a href="#">Build 1.1</a>	Q B R G
<i>Candida glabrata</i>		<a href="#">Build 1.1</a>	Q B R
<i>Cryptococcus neoformans</i>		<a href="#">Build 2.1</a>	Q B R
<i>Debaryomyces hansenii</i>		<a href="#">Build 1.1</a>	Q B R
<i>Encephalitozoon cuniculi</i>		<a href="#">Build 1.1</a>	Q B R
<i>Eremothecium gossypii</i>		<a href="#">Build 3.1</a>	Q B R
<i>Gibberella zeae</i>		<a href="#">Build 1.2</a>	Q B R
<i>Kluyveromyces lactis</i>		<a href="#">Build 1.1</a>	Q B R
<i>Magnaporthe oryzae</i>	rice blast fungus	<a href="#">Build 3.1</a>	Q B R



Entrez Genome view - Windows Internet Explorer

http://www.ncbi.nlm.nih.gov/projects/mapview/map\_search.cgi?taxid=9606

File Edit View Favorites Tools Help

Digital Differential Display Entrez Genome view

NCBI

NCBI Map Viewer

PubMed Nucleotide Protein Genome Gene Structure PopSet Taxonomy Help

Search for on chromosome(s) assembly All Find Advanced Search

Map Viewer

Map Viewer Home

Map Viewer Help

Human Maps Help

Release Notes

NCBI Resources

Genome Project

TaxPlot

Consensus CoDing Sequence (CCDS)

GRC

Human Genome Resources

NCBI Handbook

RefSeq

Trace Archive (Watson)

Trace Archive (Venter)

Trace FTP (Personal Genomics)

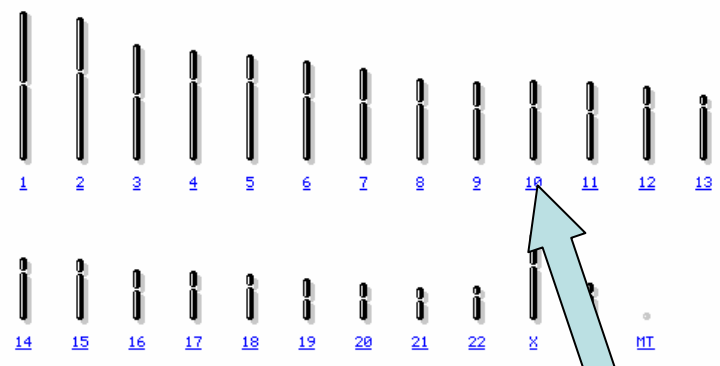
Whole Genome Association (WGA)

Organism Data in

**Homo sapiens (human) genome view**

Build 37.1 statistics Switch to previous build

BLAST search the human genome



Lineage: Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eumetazoa; Mollusca; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorhini; Catarrhini; Hominidae; Homo; Homo sapiens

**August 2009:** NCBI released an updated version of the human genome reference genome assembly and updated annotation for all available assemblies. The reference assembly update includes modifications to all chromosomes and adds 9 alternate loci to the reference assembly definition; the updated assembly, named GRCh37, was provided by the [Genome Reference Consortium \(GRC\)](#). The previous version of the reference genome assembly, [NCBI Build 36.3](#), can still be accessed for Map Viewer display and for BLAST. For additional information about changes, statistics, and the status of the CCDS project please refer to:

- [Release Notes](#)
- [Statistics](#)
- [CCDS Project](#)

The NCBI Map Viewer provides graphical displays of features on the human genome sequence assembly as well as cytogenetic, genetic, physical, and radiation hybrid maps. Extensive [documentation](#) is provided to describe the resource features and methods used, tutorials, and statistics.

Map features that can be seen along the sequence include genes, transcripts, NCBI entries (the IContig map), the BAC tiling path (the IComponent map), STGs, FISH

NCBI NCBI Map Viewer

PubMed Entrez BLAST OMIM Taxonomy Structure

Search Find Find in This View Advanced Search

[Homo sapiens \(human\) Build 37.1 \(Current\)](#) [BLAST The Human Genome](#)

Chromosome: [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [11](#) [12](#) [13](#) [14](#) [15](#) [16](#) [17](#) [18](#) [19](#) [20](#) [21](#) [22](#) [X](#) [Y](#) [MT](#)

Master Map: Genes On Sequence [Summary of Maps](#) [Map & Options](#)

Region Displayed: 0-136M bp

[Ideogram](#) [Contig](#) [Hs Unig](#) [Genes\\_seq](#)

Symbol	Links	E	Cyto	Description
<a href="#">RPL12P28</a>	<a href="#">HGNC</a> <a href="#">sv</a> <a href="#">dl</a> <a href="#">ev</a> <a href="#">mm</a>	best RefSeq	10	ribosomal protein L12 pseudogene
<a href="#">LOC100287942</a>	<a href="#">sv</a> <a href="#">dl</a> <a href="#">ev</a> <a href="#">mm</a>	mRNA	10	hypothetical protein LOC100287942
<a href="#">PRTFDC1</a>	<a href="#">OMIM</a> <a href="#">HGNC</a> <a href="#">sv</a> <a href="#">pr</a> <a href="#">dl</a> <a href="#">ev</a> <a href="#">mm</a> <a href="#">hm</a> <a href="#">sts</a> <a href="#">SNP</a>	best RefSeq	10	phosphoribosyl transferase domain
<a href="#">CCNY</a>	<a href="#">OMIM</a> <a href="#">HGNC</a> <a href="#">sv</a> <a href="#">pr</a> <a href="#">dl</a> <a href="#">ev</a> <a href="#">mm</a> <a href="#">hm</a> <a href="#">sts</a> <a href="#">SNP</a>	best RefSeq	10	cyclin Y
<a href="#">LOC100129622</a>	<a href="#">sv</a> <a href="#">pr</a> <a href="#">dl</a> <a href="#">ev</a> <a href="#">mm</a> <a href="#">hm</a> <a href="#">sts</a> <a href="#">SNP</a>	best RefSeq	10	Ras suppressor protein 1 pseudogene
<a href="#">HNRNPA3P1</a>	<a href="#">HGNC</a> <a href="#">sv</a> <a href="#">dl</a> <a href="#">ev</a> <a href="#">mm</a> <a href="#">sts</a>	best RefSeq	10	heterogeneous nuclear ribonucleoprotein
<a href="#">GDF2</a>	<a href="#">OMIM</a> <a href="#">HGNC</a> <a href="#">sv</a> <a href="#">pr</a> <a href="#">dl</a> <a href="#">ev</a> <a href="#">mm</a> <a href="#">hm</a> <a href="#">sts</a> <a href="#">SNP</a>	best RefSeq	10	growth differentiation factor 2
<a href="#">AGAP7</a>	<a href="#">HGNC</a> <a href="#">sv</a> <a href="#">pr</a> <a href="#">dl</a> <a href="#">ev</a> <a href="#">mm</a> <a href="#">hm</a> <a href="#">sts</a> <a href="#">SNP</a>	best RefSeq	10	ArfGAP with GTPase domain, ar
<a href="#">SLC9A3P</a>	<a href="#">HGNC</a> <a href="#">sv</a> <a href="#">dl</a> <a href="#">ev</a> <a href="#">mm</a>	best RefSeq	10	solute carrier family 9 (sodium/hy
<a href="#">CSTF2T</a>	<a href="#">OMIM</a> <a href="#">HGNC</a> <a href="#">sv</a> <a href="#">pr</a> <a href="#">dl</a> <a href="#">ev</a> <a href="#">mm</a> <a href="#">hm</a> <a href="#">sts</a> <a href="#">SNP</a>	best RefSeq	10q11	cleavage stimulation factor, 3' pre
<a href="#">LOC389970</a>	<a href="#">sv</a> <a href="#">dl</a> <a href="#">ev</a> <a href="#">mm</a>	best RefSeq	10	glyceraldehyde-3-phosphate dehy
<a href="#">LOC84989</a>	<a href="#">sv</a> <a href="#">pr</a> <a href="#">dl</a> <a href="#">ev</a> <a href="#">mm</a>	best RefSeq	10	hypothetical LOC84989
<a href="#">LOC283050</a>	<a href="#">sv</a> <a href="#">pr</a> <a href="#">dl</a> <a href="#">ev</a> <a href="#">mm</a> <a href="#">sts</a>	best RefSeq	10	hypothetical LOC283050
<a href="#">TPRX1P2</a>	<a href="#">HGNC</a> <a href="#">sv</a> <a href="#">dl</a> <a href="#">ev</a> <a href="#">mm</a>	best RefSeq	10q22.3	tetra-peptide repeat homeobox 1
<a href="#">C10orf57</a>	<a href="#">HGNC</a> <a href="#">sv</a> <a href="#">pr</a> <a href="#">dl</a> <a href="#">ev</a> <a href="#">mm</a> <a href="#">hm</a> <a href="#">sts</a> <a href="#">SNP</a>	best RefSeq	10	chromosome 10 open reading fra
<a href="#">HTR7</a>	<a href="#">OMIM</a> <a href="#">HGNC</a> <a href="#">sv</a> <a href="#">pr</a> <a href="#">dl</a> <a href="#">ev</a> <a href="#">mm</a> <a href="#">hm</a> <a href="#">sts</a> <a href="#">SNP</a>	best RefSeq	10q21-q24	5-hydroxytryptamine (serotonin)
<a href="#">ALDH18A1</a>	<a href="#">OMIM</a> <a href="#">HGNC</a> <a href="#">sv</a> <a href="#">pr</a> <a href="#">dl</a> <a href="#">ev</a> <a href="#">mm</a> <a href="#">hm</a> <a href="#">sts</a> <a href="#">SNP</a>	best RefSeq	10q24.3	aldehyde dehydrogenase 18 fami
<a href="#">COL17A1</a>	<a href="#">OMIM</a> <a href="#">HGNC</a> <a href="#">sv</a> <a href="#">pr</a> <a href="#">dl</a> <a href="#">ev</a> <a href="#">mm</a> <a href="#">hm</a> <a href="#">sts</a> <a href="#">SNP</a>	best RefSeq	10q24.3	collagen, type XVII, alpha 1
<a href="#">CLRN3</a>	<a href="#">HGNC</a> <a href="#">sv</a> <a href="#">pr</a> <a href="#">dl</a> <a href="#">ev</a> <a href="#">mm</a> <a href="#">hm</a> <a href="#">sts</a> <a href="#">SNP</a>	best RefSeq	10	clarin 3
<a href="#">LOC728022</a>	<a href="#">sv</a> <a href="#">pr</a> <a href="#">dl</a> <a href="#">ev</a> <a href="#">mm</a> <a href="#">hm</a>	mRNA	10	similar to double homeobox, 4

Summary of Maps:  
Map 1: Ideogram  
Region Displayed: 10pter-10qter  
Map 2: Contig

[Table View](#)

Map Viewer Home  
Map Viewer Help  
Human Maps Help  
FTP  
Data As Table View  
[Maps & Options](#)  
Compress Map  
Region Shown:  
out  
zoom  
in  
You are here:  
Ideogram  
default  
master

Map Viewer - Windows Internet Explorer

http://www.ncbi.nlm.nih.gov/projects/mapview/maps\_options.cgi?TAXID=9606&CHR=10&MAPS=ideogr,cntg-r,ugHs,genes[1.00%3A135534747.00]

Organism: **Homo sapiens** [Help](#)

Chromosome:  Region Shown:

**Available Maps:**

Org:  Assembly:

([R] before map means 'ruler set')

**Maps Displayed (left to right):**

☐ ugr is ☐ Gene

**More Options:**

☐ Show Connections ☒ Verbose Mode

Compress Map:  Auto Compress if >  px

Page Length:

Thumbnail View: ☒ default (ideogram) ☐ master

microsatellites

SNPs



NCBI NCBI Map Viewer

PubMed Entrez BLAST OMIM Taxonomy Structure

Search D10S1573 OR D10S1749 Find Find in This View Advanced Search

*Homo sapiens (human) Build 37.1 (Current)* BLAST The Human Genome

Chromosome: 1 2 3 4 5 6 7 8 9 [ 10 ] 11 12 13 14 15 16 17 18 19 20 21 22 X Y MT

Master Map: Genes On Sequence Summary of Maps Maps & Options

Region Displayed: 0-136M bp Download/View Sequence/Evidence

Ideogram	Contig	Hs Unig	Genes_seq	Symbol	Q	Accession	E	Cyto	Description
		Hs.26010		RPL12P	+	HGNC sv d ev mm	best RefSeq	10	ribosomal protein L12 pseudogen
		Hs.528300		LOC100287942	+	sv d ev mm	mRNA	10	hypothetical protein LOC100287
		Hs.460250		LOC100287942	+	OMIM HGNC sv pr d ev mm hm sts	SNP	10	phosphoribosyl transferase doma
		Hs.299055		CCNY	+	OMIM HGNC sv pr d ev mm hm sts	SNP	10	cyclin Y
		Hs.508148		LOC100287942	+	sv pr d ev mm hm	SNP	10	Ras suppressor protein 1 pseudo
		Hs.499145		HNRNPA5P1	+	HGNC sv d ev mm sts	best RefSeq	10	heterogeneous nuclear ribonucleo
		Hs.435610		GDF2	+	OMIM HGNC sv pr d ev mm hm sts	SNP	10	growth differentiation factor 2
		Hs.713531		AGAP7	+	HGNC sv pr d ev mm hm sts	SNP	10	ArfGAP with GTPase domain, ar
				SLC9A3P	+	HGNC sv d ev mm	best RefSeq	10	solute carrier family 9 (sodium/hy
				CSTF2T	+	OMIM HGNC sv pr d ev mm hm sts	SNP	10q11	cleavage stimulation factor, 3' pre
		Hs.523004		LOC389970	+	sv d ev mm	best RefSeq	10	glyceraldehyde-3-phosphate deh
		Hs.523012		LOC84989	+	sv pr d ev mm	best RefSeq	10	hypothetical LOC84989
		Hs.643896		LOC283050	+	sv pr d ev mm sts	best RefSeq	10	hypothetical LOC283050
		Hs.356927		TPRX1P2	+	HGNC sv d ev mm	best RefSeq	10q22.3	tetra-peptide repeat homeobox 1
		Hs.356794		C10orf57	+	HGNC sv pr d ev mm hm sts	SNP	10	chromosome 10 open reading fra
		Hs.530291		HTR7	+	OMIM HGNC sv pr d ev mm hm sts	SNP	10q21-q24	5-hydroxytryptamine (serotonin)
		Hs.352656		ALDH18A1	+	OMIM HGNC sv pr d ev mm hm sts	SNP	10q24.3	aldehyde dehydrogenase 18 fami
		Hs.500483		COL17A1	+	OMIM HGNC sv pr d ev mm hm sts	SNP	10q24.3	collagen, type XVII, alpha 1
		Hs.643030		CLRN3	+	HGNC sv pr d ev mm hm sts	SNP	10	clarin 3
		Hs.632918		LOC728022	+	sv pr d ev mm hm	mRNA	10	similar to double homeobox, 4
		Hs.500721							
		Hs.500756							
		Hs.558396							
		Hs.597496							
		Hs.500842							
		Hs.153961							
		Hs.501012							
		Hs.501135							
		Hs.523302							
		Hs.523332							
		Hs.656386							

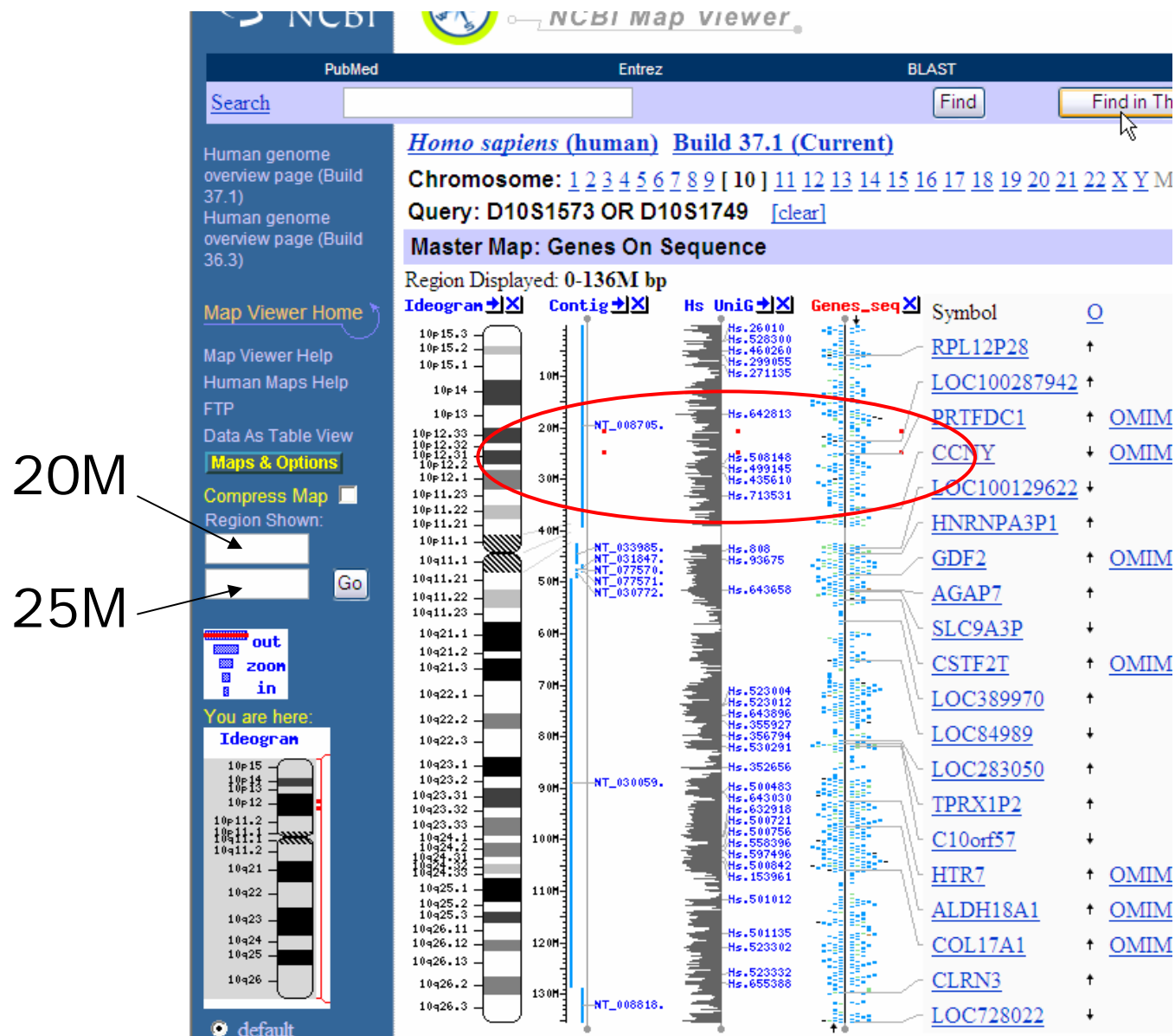
Summary of Maps:  
Map 1: Ideogram  
Region Displayed: 10pter-10qter  
Map 2: Contig

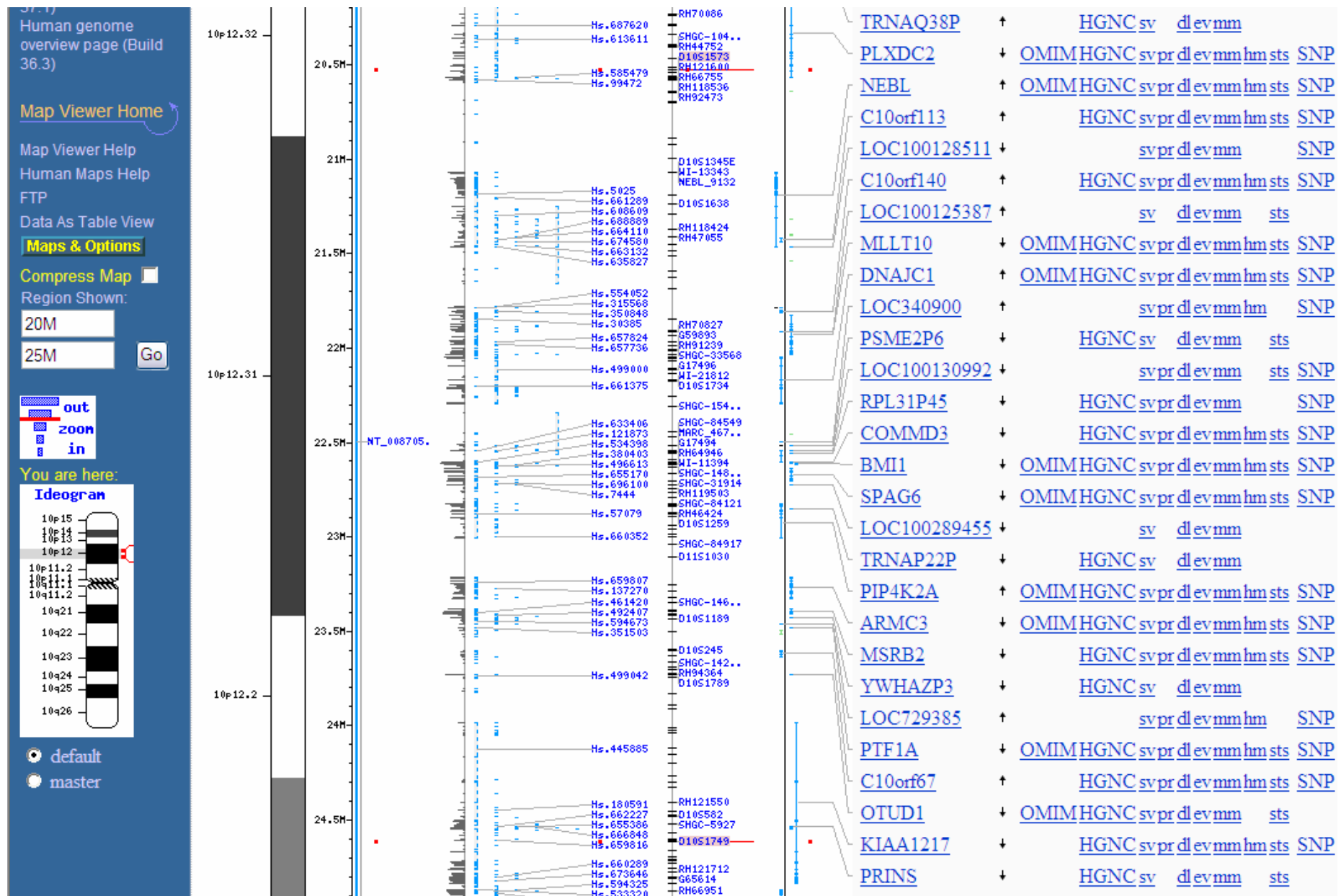
Table View

absolutely not this!!

this







Map Viewer - Windows Internet Explorer

http://www.ncbi.nlm.nih.gov/projects/mapview/maps.cgi?TAXID=9606&CHR=10&MAPS=ideogr,cntg-r,ughs,sts,genes[20000000.00%3A25000000.00]&query=D1051573%20%

File Edit View Favorites Tools Help

Map Viewer

Human genome overview page (Build 37.1)  
Human genome overview page (Build 36.3)

[Map Viewer Home](#)

[Map Viewer Help](#)  
[Human Maps Help](#)  
[FTP](#)  
[Data As Table View](#)  
**[Maps & Options](#)**

**Compress Map** ☐

Region Shown:  
20M  
25M

**out**  
**zoom**  
**in**

You are here:  
**Ideogram**

10p15  
10p13  
10p12  
10p11.2  
10p11.1  
10p11.2  
10q21  
10q22  
10q23  
10q24  
10q25  
10q26

☒ default  
☐ master

Human genome overview page (Build 37.1)  
Human genome overview page (Build 36.3)

23H  
23.5H  
24H  
24.5H

10p12.2

Hs.57079  
Hs.660352  
Hs.659807  
Hs.137279  
Hs.461420  
Hs.492407  
Hs.594573  
Hs.351503  
Hs.445885  
Hs.180591  
Hs.662227  
Hs.655386  
Hs.666048  
Hs.659816  
Hs.660289  
Hs.673646  
Hs.594325  
Hs.533320  
Hs.524195

RH46424  
D1051259  
SHGC-84917  
D1151030  
SHGC-146..  
D1051189  
D105245  
SHGC-142..  
RH94364  
D1051789  
RH121550  
D105582  
SHGC-5927  
D1051749  
RH121712  
G05614  
RH60951

[LOC100289455](#) + [sv](#) [dl](#) [ev](#) [mm](#) mRNA 10 hypothe  
[TRNAP22P](#) + [HGNC](#) [sv](#) [dl](#) [ev](#) [mm](#) best RefSeq 10 transfe  
[PIP4K2A](#) + [OMIM](#) [HGNC](#) [sv](#) [pr](#) [dl](#) [ev](#) [mm](#) [hm](#) [sts](#) [SNP](#) best RefSeq 10 phosph  
[ARMC3](#) + [OMIM](#) [HGNC](#) [sv](#) [pr](#) [dl](#) [ev](#) [mm](#) [hm](#) [sts](#) [SNP](#) best RefSeq 10p12.31 armadi  
[MSRB2](#) + [HGNC](#) [sv](#) [pr](#) [dl](#) [ev](#) [mm](#) [hm](#) [sts](#) [SNP](#) best RefSeq 10p12 methio  
[YWHAZP3](#) + [HGNC](#) [sv](#) [dl](#) [ev](#) [mm](#) best RefSeq 10 tyrosin  
[LOC729385](#) + [sv](#) [pr](#) [dl](#) [ev](#) [mm](#) [hm](#) [sts](#) [SNP](#) mRNA 10 hypothe  
[PTF1A](#) + [OMIM](#) [HGNC](#) [sv](#) [pr](#) [dl](#) [ev](#) [mm](#) [hm](#) [sts](#) [SNP](#) best RefSeq 10 pancre  
[C10orf67](#) + [HGNC](#) [sv](#) [pr](#) [dl](#) [ev](#) [mm](#) [hm](#) [sts](#) [SNP](#) best RefSeq 10 chrom  
[OTUD1](#) + [OMIM](#) [HGNC](#) [sv](#) [pr](#) [dl](#) [ev](#) [mm](#) [sts](#) best RefSeq 10 OTU c  
[KIAA1217](#) + [HGNC](#) [sv](#) [pr](#) [dl](#) [ev](#) [mm](#) [hm](#) [sts](#) [SNP](#) best RefSeq 10 KIAA  
[PRINS](#) + [HGNC](#) [sv](#) [dl](#) [ev](#) [mm](#) [sts](#) best RefSeq 10p12.1 psorias  
[ARHGAP21](#) + [OMIM](#) [HGNC](#) [sv](#) [pr](#) [dl](#) [ev](#) [mm](#) [hm](#) [sts](#) [SNP](#) best RefSeq 10 Rho G

**Summary of Maps:**

**Map 1: Ideogram**  
Region Displayed: 10p12.32-10p12.1  
[Table View](#)

**Map 2: Contig**  
Region Displayed: 20M-25M bp  
Total Contigs On Chromosome: 8  
Contigs Labeled: 1 Total Contigs in Region: 1  
[Download/View Sequence/Evidence](#)

**Map 3: Homo sapiens UniGene Clusters**  
Region Displayed: 20M-25M bp  
Total Transcript alignments On Chromosome: 6607  
UniGene Clusters Labeled: 50 Total Transcript alignments in Region: 206  
Histogram Data: Tick Width=6,702bp/pixel, Max Height=246 transcripts (logarithmic scale)  
[Table View](#)

**Map 4: STS**  
Region Displayed: 20M-25M bp  
Total STSs On Chromosome: 8594  
STSs Labeled: 50 Total STSs in Region: 213  
[Download/View Sequence/Evidence](#)

**Map 5: Genes On Sequence**  
Region Displayed: 20M-25M bp  
Total Genes On Chromosome: 1305  
Genes Labeled: 30 Total Genes in Region: 40  
[Table View](#) [Download/View Sequence/Evidence](#)

[Disclaimer](#) | [Write to the Help Desk](#)  
[NCBI](#) | [NLM](#) | [NIH](#)



Map Viewer - Windows Internet Explorer

http://www.ncbi.nlm.nih.gov/projects/mapview/maps.cgi?TAXID=9606&CHR=10&MAPS=ideogr,cntg-r,ugHs,sts,genes[20000000.00%3A25000000.00]&QUERY=D1051573+++

File Edit View Favorites Tools Help

Map Viewer

Genes On Sequence All Sequence Maps [next](#)

Region Displayed: 20M-25M bp

Total Genes On Chromosome: 1305

Genes in Region: 40

[Download/View Sequence Evidence](#) [Download Data](#)

start	stop	Symbol	O	Links	E	Cyto	Description
19422755	20023407	<a href="#">C10orf112</a>	+	<a href="#">HGNC</a> <a href="#">sv pr dl ev mm hm sts</a> <a href="#">SNP</a>	mRNA	10	chromosome 10 open reading frame 112
20036608	20036682	<a href="#">TRNAQ38P</a>	-	<a href="#">HGNC</a> <a href="#">sv</a> <a href="#">dl ev mm</a>	best RefSeq	10	transfer RNA glutamine 38 (anticodon CUG) pseudogene
20105372	20569115	<a href="#">PLXDC2</a>	+	<a href="#">OMIM</a> <a href="#">HGNC</a> <a href="#">sv pr dl ev mm hm sts</a> <a href="#">SNP</a>	best RefSeq	10	plexin domain containing 2
20639169	20640038	<a href="#">LOC100133024</a>	+	<a href="#">sv</a> <a href="#">dl ev mm</a> <a href="#">sts</a>	protein	10	similar to adenosylmethionine decarboxylase 1
21068902	21463116	<a href="#">NEBL</a>	-	<a href="#">OMIM</a> <a href="#">HGNC</a> <a href="#">sv pr dl ev mm hm sts</a> <a href="#">SNP</a>	best RefSeq	10p12	nebulin
21317928	21319745	<a href="#">LOC100129337</a>	+	<a href="#">sv</a> <a href="#">dl ev mm</a>	protein	10	hypothetical LOC100129337
21399611	21402484	<a href="#">LOC100129615</a>	+	<a href="#">sv</a> <a href="#">dl ev mm</a>	protein	10	similar to nucleophosmin 1
21414692	21435427	<a href="#">C10orf113</a>	-	<a href="#">HGNC</a> <a href="#">sv pr dl ev mm hm sts</a> <a href="#">SNP</a>	best RefSeq	10	chromosome 10 open reading frame 113
21462202	21463850	<a href="#">LOC100128511</a>	+	<a href="#">sv pr dl ev mm</a> <a href="#">SNP</a>	mRNA	10	hypothetical LOC100128511
21536681	21537001	<a href="#">LOC100289386</a>	+	<a href="#">sv</a> <a href="#">dl ev mm</a>	protein	10	hypothetical LOC100289386
21783421	21786213	<a href="#">C10orf114</a>	-	<a href="#">HGNC</a> <a href="#">sv pr dl ev mm</a> <a href="#">SNP</a>	best RefSeq	10	chromosome 10 open reading frame 114
21785491	21785570	<a href="#">MIR1915</a>	-	<a href="#">sv</a> <a href="#">dl ev mm</a>	external	10	microRNA 1915
21802407	21814611	<a href="#">C10orf140</a>	-	<a href="#">HGNC</a> <a href="#">sv pr dl ev mm hm sts</a> <a href="#">SNP</a>	best RefSeq	10	chromosome 10 open reading frame 140
21823102	22032554	<a href="#">MLLT10</a>	+	<a href="#">OMIM</a> <a href="#">HGNC</a> <a href="#">sv pr dl ev mm hm sts</a> <a href="#">SNP</a>	best RefSeq	10p12	myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila); translocated
21912854	21915285	<a href="#">LOC100125387</a>	-	<a href="#">sv</a> <a href="#">dl ev mm</a> <a href="#">sts</a>	best RefSeq	10	heterogeneous nuclear ribonucleoprotein R pseudogene
22045477	22292650	<a href="#">DNAJC1</a>	-	<a href="#">OMIM</a> <a href="#">HGNC</a> <a href="#">sv pr dl ev mm hm sts</a> <a href="#">SNP</a>	best RefSeq	10	DnaJ (Hsp40) homolog, subfamily C, member 1
22453382	22456165	<a href="#">LOC645220</a>	+	<a href="#">sv</a> <a href="#">dl ev mm</a>	protein	10	similar to adiponectin receptor 1
22497812	22498912	<a href="#">LOC340900</a>	-	<a href="#">sv pr dl ev mm hm</a> <a href="#">SNP</a>	protein	10	similar to hCG1652542
22514626	22515000	<a href="#">PSME2P6</a>	+	<a href="#">HGNC</a> <a href="#">sv</a> <a href="#">dl ev mm</a> <a href="#">sts</a>	best RefSeq	10p12	proteasome activator subunit 2 pseudogene 6
22518438	22518511	<a href="#">TRNAN-GUU</a>	-	<a href="#">sv</a> <a href="#">dl ev mm</a> <a href="#">sts</a>	tRNA scan SE	10	transfer RNA asparagine (anticodon GUU)
22541001	22543553	<a href="#">LOC100130992</a>	+	<a href="#">sv pr dl ev mm</a> <a href="#">sts</a> <a href="#">SNP</a>	mRNA	10	similar to hCG2017625
22542182	22606590	<a href="#">LOC100287942</a>	-	<a href="#">sv</a> <a href="#">dl ev mm</a>	mRNA	10	hypothetical protein LOC100287942
22555919	22556134	<a href="#">RPL31P45</a>	+	<a href="#">HGNC</a> <a href="#">sv pr dl ev mm</a> <a href="#">SNP</a>	best RefSeq	10	ribosomal protein L31 pseudogene 45
22605299	22609237	<a href="#">COMMD3</a>	+	<a href="#">HGNC</a> <a href="#">sv pr dl ev mm hm sts</a> <a href="#">SNP</a>	best RefSeq	10pter-422.1	COMM domain containing 3
22610140	22620188	<a href="#">BMI1</a>	+	<a href="#">OMIM</a> <a href="#">HGNC</a> <a href="#">sv pr dl ev mm hm sts</a> <a href="#">SNP</a>	best RefSeq	10p11.23	BMI1 polycomb ring finger oncogene
22634399	22706539	<a href="#">SPAG6</a>	+	<a href="#">OMIM</a> <a href="#">HGNC</a> <a href="#">sv pr dl ev mm hm sts</a> <a href="#">SNP</a>	best RefSeq	10p12.2	sperm associated antigen 6

# Position-independent gene identification

## [3. The “candidate gene” approach]

- The choice of candidates is based on:
  1. Tissue specificity and abundance
  2. Chromosomal location
  3. Sequence information and length
  4. Similarity to other known disease-causing genes
  5. Other characteristics (gene conservation, etc.)

Ensembl Genome Browser - Windows Internet Explorer

http://www.ensembl.org/index.html

File Edit View Favorites Tools Help

Ensembl Genome Browser

**e!Ensembl**  
Home

Login / Register | BLAST/BLAT | BioMart | Docs & FAQs | Mirrors

Search: All species for BMI1 Go

rat X:100000..200000 or insulin

**Browse a genome**

The Ensembl project provides a comprehensive set of genome browser tools for vertebrates and other species available online. Click on a link to go to the species' home page.

**Popular genomes**

- Human (GRCh37)
- Mouse (NCBIM37)
- Zebrafish (Zv8)

**All genomes**

-- Select a species --

[View full list of all Ensembl species](#)

Other pre-build species are available in [Ensembl Pre! →](#)

**New to Ensembl?**

Did you know you can:

- [Learn how to use Ensembl](#) with our video tutorials and walk-throughs
- [Add custom tracks](#) using our new Control Panel
- [Upload your own data](#) and save it to your Ensembl account
- [Search for a DNA or protein sequence](#) using BLAST or BLAT
- [Fetch only the data you want](#) from our public database, using the Ensembl Perl API
- [Download our databases via FTP](#) in FASTA, MySQL and other formats
- [Mine Ensembl with BioMart](#) and export sequences or tables in text, html, or Excel format

Still got questions? [Try our FAQs](#)

**Did you know...?**

Browsing Ensembl from North America or the Pacific Rim? Try our mirror site, <http://uswest.ensembl.org> for faster access.

**What's New in Release 56 (15 September 2009)**

- **New species - pig** (Pig)
- **New species - marmoset** (Marmoset)
- **New rat gene set** (Rat)
- **Multi-species views** (all species)
- **New Regulatory Feature Panel** (Human, Mouse)

type in

choose

Ensembl genome browser 56: H.sapiens - Gene summary - Gene: BMI1 (ENSG00000168283) - Windows Internet Explorer

http://www.ensembl.org/Homo\_sapiens/Gene/Summary?g=ENSG00000168283

File Edit View Favorites Tools Help

Ensembl genome browser 56: H.sapiens - Gene summ...

**Synonyms** PCGF4, RNF51 [To view all Ensembl genes linked to the name [click here.](#)]

**CCDS** This gene is a member of the Human CCDS set: [CCDS7138](#)

**Gene type** Known protein coding

**Prediction Method** Gene containing both Ensembl genebuild transcripts and [Havana](#) manual curation, see [article.](#)

**Alternative genes** This Known protein coding entry corresponds to the following database identifiers:  
Havana Gene: [OTTHUMG00000017807](#) [[view all locations](#)]

**Transcripts**

Ensembl/Havana g...

22.60 Mb 22.61 Mb 22.61 Mb 22.61 Mb 22.62 Mb 22.62 Mb 22.62 Mb

35.03 Kb Forward strand

BMI1-001 > Known protein coding Ensembl/Havana merge gene

BMI1-002 > Known protein coding Havana gene

BMI1-004 > Known protein coding Havana gene

BMI1-201 > Known protein coding Ensembl gene

COMMD3-001 > Known protein coding Ensembl/Havana merge gene

COMMD3-002 > Known protein coding Havana gene

COMMD3-003 > Known protein coding Havana gene

COMMD3-005 > Known protein coding Havana gene

COMMD3-004 > Known protein coding Havana gene

COMMD3-006 > Known protein coding Havana gene

COMMD3-007 > Known protein coding Havana gene

COMMD3-008 > Known protein coding Havana gene

COMMD3-009 > Known protein coding Havana gene

COMMD3-010 > Known protein coding Havana gene

COMMD3-011 > Known protein coding Havana gene

**GNC (curated): BMI1-001**

Transcript: [ENST00000375663](#)

Gene: [ENSG00000168283](#)

Protein product: [ENSP000003651](#)

Location: **Chromosome 10:**  
22,610,140-22,620,413

Gene type: Known Protein coding

Strand: Forward

Base pairs: 3,428

Amino acids: 326

Analysis: Ensembl/Havana merge gene

Gene containing both Ensembl genebuild transcripts and [Havana](#) manual curation, see [article.](#)

Ensembl genome browser 56: H.sapiens - Transcript summary - Transcript: BMI1-001 (ENST00000376663 - Windows Internet Explorer

http://www.ensembl.org/Homo\_sapiens/Transcript/Summary?db=core;g=ENSG00000168283;r=10:22605381-22620413;t=ENST00000376663

File Edit View Favorites Tools Help

Ensembl genome browser 56: H.sapiens - Transcript s...

**e!Ensembl**

Home > Human [GRCh37]

Login / Register | BLAST/BLAT | BioMart | Docs & FAQs | Mirrors

Location: 10:22,605,381-22,620,413 Gene: BMI1 Transcript: BMI1-001

**Transcript-based displays**

- Transcript summary
- Supporting evidence (80)
- Sequence
  - Exons (10)
  - cDNA
  - Protein
- External References
  - General identifiers (1)
  - Oligo probes (20)
  - Gene ontology (10)
- Genetic Variation
  - Population comparison
  - Comparison image
- Protein Information
  - Protein summary
  - Domains & features (2)
  - Variations (2)
- External Data
  - Personal annotation
- ID History
  - Transcript history
  - Protein history

• Configure this page  
• Manage your data  
• Export data  
• Bookmark this page

**Transcript: BMI1-001 (ENST00000376663)**

Polycomb complex protein BMI-1 (Polycomb group RING finger protein 4)(RING finger protein 51) [Source:UniProtKB/Swiss-Prot;Acc:P35226]

**Location** [Chromosome 10: 22,610,140-22,620,413](#) forward strand.

**Gene** This transcript is a product of gene [ENSG00000168283](#) - There are 11 transcripts in this gene: [hide transcripts](#)

Name	Transcript ID	Protein ID	Description
BMI1-001	<a href="#">ENST00000376663</a>	<a href="#">ENSP00000365851</a>	protein_coding
BMI1-002	<a href="#">ENST00000442508</a>	<a href="#">ENSP00000397912</a>	protein_coding
BMI1-003	<a href="#">ENST00000456675</a>	<a href="#">ENSP00000401773</a>	protein_coding
BMI1-004	<a href="#">ENST00000417470</a>	<a href="#">ENSP00000398759</a>	protein_coding
BMI1-005	<a href="#">ENST00000443519</a>	<a href="#">ENSP00000390768</a>	protein_coding
BMI1-006	<a href="#">ENST00000496174</a>	No protein product	processed_transcript
BMI1-007	<a href="#">ENST00000490311</a>	No protein product	processed_transcript
BMI1-008	<a href="#">ENST00000416820</a>	<a href="#">ENSP00000399220</a>	protein_coding
BMI1-201	<a href="#">ENST00000376691</a>	<a href="#">ENSP00000365881</a>	protein_coding
BMI1-202	<a href="#">ENST00000376719</a>	<a href="#">ENSP00000365909</a>	protein_coding
BMI1-203	<a href="#">ENST00000415729</a>	<a href="#">ENSP00000389191</a>	protein_coding

**Transcript and Gene level displays**

In Ensembl a gene is made up of one or more transcripts. Views in Ensembl are separated into Gene based views and Transcript based views according to which level the information is more appropriately associated. This view is a transcript level view. To flip between the two sets of views you can click on the Gene and Transcript tabs in the menu bar at the top of the page.

**Transcript summary** [help](#) [Supporting evidence »](#)

10.27 Kb

Forward strand

[Export Image](#)

**Statistics**

**CCDS** Exons: 10 Transcript length: 3,428 bps Translation length: 326 residues  
This transcript is a member of the Human CCDS set: [CCDS7138](#)

**Type** Known protein coding

**Prediction Method** Transcript where the Ensembl genebuild transcript and the [Vega](#) manual annotation have the same sequence, for every base pair. See [article](#).

**Alternative transcripts** This Ensembl/Havana merge gene entry corresponds to the following database identifiers:

Ensembl genome browser 56: H.sapiens - Exons - Transcript: BMI1-001 (ENST00000376663) - Windows Internet Explorer

http://www.ensembl.org/Homo\_sapiens/Transcript/Exons?db=core;g=ENSG00000168283;r=10:22605381-22620413;t=ENST00000376663

File Edit View Favorites Tools Help

Ensembl genome browser 56: H.sapiens - Exons - Tra...

Manage your data  
Export data  
Bookmark this page

« Supporting evidence Exons [help](#) cDNA sequence »

No.	Exon / Intron	Start	End	Start Phase	End Phase	Length	Sequence
	5' upstream sequence						.....ccccgctgcacgcacacacggcgcccccgccccgccccgctccccca
1	<a href="#">ENSE00001251863</a>	22,610,140	22,610,625	-	-	486	CAGCAACTATGAAATAATCGTAGTATGAGAGGCAGAGATCGGGGCGAGACAATGGGGATG TGGGCGCGGGAGCCCCGTTCCGGCTTAGCAGCACCTCCAGCCCCGAGAAATAAACCGA TCGCGCCCCCTCCGCGCGCGCCCTCCCCGAGTGCAGGAGCGGGAGGAGCGCGCGCGCC GAGGAGGAGGAGGAGGAGGCCCCGAGGAGGAGGCGTTGGAGGTCGAGGCGGAGGCGGAG GAGGAGGAGGCGGAGGCGCGGAGGAGGCGGAGGCGCGGAGCAGGAGGAGGCGCGCGG AGGCGGCATGAGACGAGCGTGGCGGCGCGGCTGCTCGGGGCGCGCTGGTTGCCATTG ACAGCGGCGTCTGCAGCTCGCTTCAAGATGGCCGCTTGGCTCGCATTCAITTTCTGCTGA ACGACTTTTAACTTTTCATTGTCTTTCCGCGCGCTTCGATCGCCTCGCGCGGCTGCTCT TTCCGG
	Intron 1-2	22,610,626	22,615,359			4,734	gtacgtaggaggcgaggcgccccg.....tgccattatttctgtgtcttgcag
2	<a href="#">ENSE00001149872</a>	22,615,360	22,615,490	-	1	131	GATTTTTATCAAGCAGAAATGCATCGAACACGAGAATCAAGATCACTGAGCTAAATCC CCACCTGATGTGTGTGCTTTGTGGAGGGTACTTCATTGATGCCACAACCATATAAGAATG TCTACATTCTCT
	Intron 2-3	22,615,491	22,615,818			328	gtaagtaccgagcttttagctctctt.....aaaacttcacatgttctacttctag
3	<a href="#">ENSE00001628582</a>	22,615,819	22,615,915	1	2	97	TCTGTAAACGTTGATTGTTTCGTTACCTGGAGACCAGCAAGTATTGTCTTATTGTGATG TCCAAGTTTCAAGACCAAGCACTACTGAATATAAG
	Intron 3-4	22,615,916	22,616,523			608	gtaggaactgttgaaattccttgt.....gtttctatttttaattatttttcag
4	<a href="#">ENSE00001137647</a>	22,616,524	22,616,579	2	1	56	GTCAGATAAACTCTCCAAGATATTGTATACAAATTAGTTCAGGGCTTTTCAAAA
	Intron 4-5	22,616,580	22,616,670			91	gtgagtaacttgcttagaaaatgaa.....gtttacatctttttcccccattcag
5	<a href="#">ENSE00001904285</a>	22,616,671	22,616,721	1	1	51	ATGAAATGAAGAGAAGAAGGGATTTTATGCAGCTCATCTTCTGCTGATG
	Intron 5-6	22,616,722	22,616,878			157	gtaaaccttttaggggagggaagac.....attacatttcactaatatcggtatag
6	<a href="#">ENSE00001948815</a>	22,616,879	22,616,987	1	2	109	CTGCCAATGGCTCTAATGAAGATAGAGGAGAGGTTGCAGATGAAGATAAGAGAATTATAA CTGATGATGAGATAATAAGCTTATCCATTGAATTCTTTGACCAGAACAG
	Intron 6-7	22,616,988	22,617,062			75	gtaaaatcttttaggcaatttattt.....cactaataaattttctctttattag
7	<a href="#">ENSE00001705310</a>	22,617,063	22,617,108	2	0	46	ATTGGATCGGAAAGTAAACAAAGACAAAGAGAAATCTAAGGAGGAG
	Intron 7-8	22,617,109	22,617,528			420	gtatgtttcatgtttacaaaaacata.....gttacttttctaaatgtacttttag
8	<a href="#">ENSE00001594852</a>	22,617,529	22,617,627	0	0	99	GTGAATGATAAAGATACTTACGATGCCAGCAGCAATGACTGTGATGCACTTAAGAAAG TTTCTCAGAAGTAAATGGACATACCTAATACTTTCCAG
	Intron 8-9	22,617,628	22,617,976			349	gtatctacttttatattcttcttgc.....acttaataaaaaatatttttctactag
9	<a href="#">ENSE00001777532</a>	22,617,977	22,618,057	0	0	81	ATTGATGTCATGTATGAGGAGGAACCTTTAAAGGATTATTATACACTAATGGATATTGCC TACATTTATACCTGGAGAAGG



TTCCGGGATTTTTTATCAAGCAGAAATGCATCGAACAACGAGAATCAAGATCACTGAGCT  
 .....ATGCATCGAACAACGAGAATCAAGATCACTGAGCT  
 .....-M--H--R--T--T--R--I--K--I--T--E--L

AAATCCCCACCTGATGTGTGTGCTTTGTGGAGGGTACTTCATTGATGCCACAACCATAAT  
 AAATCCCCACCTGATGTGTGTGCTTTGTGGAGGGTACTTCATTGATGCCACAACCATAAT  
 --N--P--H--L--M--C--V--L--C--G--G--Y--F--I--D--A--T--T--I--I

AGAATGTCTACATTCCCTTCTGTAAAACGTGTATTGTTTCGTTACCTGGAGACCAGCAAGTA  
 AGAATGTCTACATTCCCTTCTGTAAAACGTGTATTGTTTCGTTACCTGGAGACCAGCAAGTA  
 --E--C--L--H--S--F--C--K--T--C--I--V--R--Y--L--E--T--S--K--Y

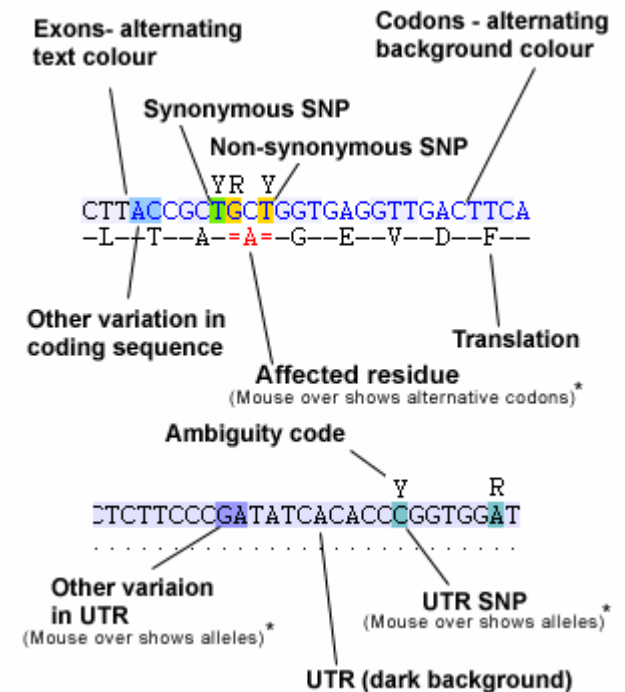
TTGTCCTATTTGTGATGTCCAAGTTCACAAGACCAGACCACTACTGAATATAAGGTCAGA  
 TTGTCCTATTTGTGATGTCCAAGTTCACAAGACCAGACCACTACTGAATATAAGGTCAGA  
 --C--P--I--C--D--V--Q--V--H--K--T--R--P--L--L--N--I--R--S--D

TAAAACTCTCCAAGATATTGTATACAAATTAGTTCAGGGCTTTTCAAAAATGAAATGAA  
 TAAAACTCTCCAAGATATTGTATACAAATTAGTTCAGGGCTTTTCAAAAATGAAATGAA  
 --K--T--L--Q--D--I--V--Y--K--L--V--P--G--L--F--K--N--E--M--K

GAGAAGAAGGGATTTTTATGCAGCTCATCCTTCTGCTGATGCTGCCAATGGCTCTAATGA  
 GAGAAGAAGGGATTTTTATGCAGCTCATCCTTCTGCTGATGCTGCCAATGGCTCTAATGA  
 --R--R--R--D--F--Y--A--A--H--P--S--A--D--A--A--N--G--S--N--E

AGATAGAGGAGAGGTTGCAGATGAAGATAAGAGAATTATAACTGATGATGAGATAATAAG  
 AGATAGAGGAGAGGTTGCAGATGAAGATAAGAGAATTATAACTGATGATGAGATAATAAG  
 --D--R--G--E--V--A--D--E--D--K--R--I--I--T--D--D--E--I--I--S

CTTATCCATTGAATTCCTTTGACCAGAACAGATTGGATCGGAAAAGTAAACAAAGACAAAGA  
 CTTATCCATTGAATTCCTTTGACCAGAACAGATTGGATCGGAAAAGTAAACAAAGACAAAGA  
 --L--S--I--E--F--F--D--O--N--R--L--D--R--K--V--N--K--D--K--E



Ensembl genome browser 56: H.sapiens - Gene summary - Gene: BMI1 (ENSG00000168283) - Windows Internet Explorer

http://www.ensembl.org/Homo\_sapiens/Gene/Summary?db=core;g=ENSG00000168283;r=10:22605381-22620413;t=ENST00000376663

File Edit View Favorites Tools Help

Ensembl genome browser 56: H.sapiens - Gene summ...

**e!Ensembl**

Home > Human [GRCh37]

Login / Register | BLAST/BLAT | BioMart | Docs & FAQs | Mirrors

Location: 10:22,605,381-22,620,413 Gene: **BMI1** Transcript: BMI1-001

**Gene-based displays**

- Gene summary
- Splice variants (11)
- Supporting evidence
- Sequences
- External features (3)
- Regulation
- Comparative genomics
  - Genomic alignments (5)
  - Gene Tree (image)
  - Gene Tree (text)
  - Gene Tree (alignment)
- Orthologues (42)
- Paralogues (5)
- Protein families (3)
- Genetic Variation
  - Variation Table
  - Variation Image
- External Data
  - Personal annotation
- ID History
  - Gene history

• Configure this page

• Manage your data

• Export data

• Bookmark this page

**Gene: BMI1 (ENSG00000168283)**

Polycarb complex protein BMI-1 (Polycarb group RING finger protein 4)(RING finger protein 51) [Source: UniProtKB/Swiss-Prot P35226](#)

**Location** [Chromosome 10: 22,605,381-22,620,413](#) forward strand.

**Transcripts** There are 11 transcripts in this gene: [hide transcripts](#)

Name	Transcript ID	Protein ID	Description
BMI1-001	<a href="#">ENST00000376663</a>	<a href="#">ENSP00000365851</a>	protein_coding
BMI1-002	<a href="#">ENST00000442508</a>	<a href="#">ENSP00000397912</a>	protein_coding
BMI1-003	<a href="#">ENST00000456675</a>	<a href="#">ENSP00000401773</a>	protein_coding
BMI1-004	<a href="#">ENST00000417470</a>	<a href="#">ENSP00000398759</a>	protein_coding
BMI1-005	<a href="#">ENST00000443519</a>	<a href="#">ENSP00000390768</a>	protein_coding
BMI1-006	<a href="#">ENST00000496174</a>	No protein product	processed_transcript
BMI1-007	<a href="#">ENST00000490311</a>	No protein product	processed_transcript
BMI1-008	<a href="#">ENST00000416820</a>	<a href="#">ENSP00000399220</a>	protein_coding
BMI1-201	<a href="#">ENST00000376691</a>	<a href="#">ENSP00000365881</a>	protein_coding
BMI1-202	<a href="#">ENST00000376719</a>	<a href="#">ENSP00000365909</a>	protein_coding
BMI1-203	<a href="#">ENST00000415729</a>	<a href="#">ENSP00000389191</a>	protein_coding

**Transcript and Gene level displays**

In Ensembl a gene is made up of one or more transcripts. We provide displays at two levels:

- Transcript views which provide information specific to an individual transcript such as the cDNA and CDS sequences and protein domain annotation.
- Gene views which provide displays for data associated at the gene level such as orthologues and paralogues, regulatory regions and splice variants.

This view is a gene level view. To access the transcript level displays select a Transcript ID in the table above and then navigate to the information you want using the menu at the left hand side of the page. To return to viewing gene level information click on the Gene tab in the menu bar at the top of the page.



Ensembl genome browser 56: H.sapiens - Marked-up sequence - Gene: BMI1 (ENSG00000168283) - Windows Internet Explorer

http://www.ensembl.org/Homo\_sapiens/Gene/Sequence?db=core;g=ENSG00000168283;r=10:22605381-22620413;t=ENST00000376663;time=1254487011843.843

File Edit View Favorites Tools Help

Ensembl genome browser 56: H.sapiens - Marked-up ...

THIS STYLE: Location of SNPs  
THIS STYLE: Location of inserts  
THIS STYLE: Location of deletes

```
>chromosome:GRCh37:10:22604781:22621013:1
GCAAGAATCACATAATTAGATGTCTTACAGTCTCATAAAAGGGCAGTGCTGCGAGGTTAA
GCTGGTGAGCTAAAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC
CTGAGATCGGAGCTTCCCGCAGGATTGCGCGAGAGCGAGGAGCGGGGAGGACGCCGTTTG
CGGGAGGCGGGGCGGGGATCCCGACACCGTCCCGCAGTCTCCCGGGGTCTCTCAGGGAA
CCGAGCTGGTGGCGTCCGAGGTAAGCGCCGAACCAAGGAGAAAGCGCCGAGCGCGTCCA
CAAAACGCCCGAGGCGGCCCTTCTCGCCAAGCTCGGCGAGTCCCGCGGCTGCGCTGGA
CAAGCGAGAGCTGGACGGAACGGTGGGAAGCCGTAGGCGCGGCTGGGGGCGCGGGACC
CGCGCGCGGGCGCGGAGGCGGGCGGGGAGGCCCCGCAAGCCGCGCGCGGCCAGCCGCC
CGAGCCGCGTGCAGCGCGCGCGCCGATCAGCTGAGCGCGCGCGCGTGTACGTGGTGTGC
GTGTCGAAGGTCACGGCGCGCTCACAATGGAGCTCTCGGAGTCTGTGCAGAAAGGCTTCC
AGATGCTGGCGGATCCCAGTCTTCGACTCCAACGCCTTCACGCTTCTCTCCGGGCGG
CATTCCAGAGTCTGTGGACGCCAGGCGGACGAGGCGGTGTAGTAAGCCGCTCGGCT
CGGCTCGGAGCTGGATCCGCCGGGTGGAGGGGCTCGGAGAAGAGGAGCGCGCGGAGC
GAGAAGGGGAAAGCCCCGGGAAGAGGAAGTCTCCGGGCTTTGCCCTTCGCTCCGGACGGA
GTGTTGGGACGTGGCTCTGCCAGGACTCGGAGCAGACTCTGAGCATTGTTTTGCCCCCA
CTTGGCACCTGCATAGAGCCAACCTCCACGTTCTGGCCCGGTTTCCCTCCAGTCTCTTT
GTTCTTTGTAAAGCCGGTGTTCAGAGGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG
CCTTGTCTGAGGCTGCGGCTGTTTGTACAGTTGGACTTCATCTGGGGAGCTCCCTGCAAT
TCAAGCTTCCAGGGCTAGGCTAGCAAGGTTTAAAGGTTCTGTTTCCGAAGTAATTACTAG
GTCAGACAGTGTGACATTTGCCAATCTAGAGCATACCTCAAGTTGCTATGATCACTCCA
CGGCTCTAGTTCAGCATGTTTACACAGGTTCTACGCTAGAACCTTAAACTATGATGCAC
AATACTTTGTTCCGCCCGAGTTACCGAGGATGCGAGTGTCTTTAAGCGATCCGGTCAGTG
ACTGCGGAGGTGCAGTGTGAGTATGGATTGAAGACATAATTTAAGACCGATCGCTTTTAG
ATGATTTTCTTAAATATAAATATCCCAACAAATCAAATTTTGTGTTTTTCTGTAGTTT
CATGAACATAACACTTGTACAAAAAGTATGAAAGTGGAGTTCTTAAACGAAAAATAGTAA
CCACCACCTGGGTTAATCTCAAAAAAAGTACTGTCAAACCTGAAAGTACTAATATA
ATTTTGAATTTGTATCTTATATTCTGAACATGCCCCACCTTCTACCCCTTGATGTA
TTTGTTTATGGAGCCGGTTACCTAAACTCATCATTTTAGGATGGAAGTTTATTCTAAA
ATTGACACAATGTAATAATTTATCTTGTGTTTCTTGAATTTACTCCAGTGTAGAAAT
GACCATTGCATTGTTATATTTTCTTTGGGGGCTATAATATTTTAAATAGTAAAAATATT
AAGGCAAAATAAGTTAATTGAAGCTAGTAATAGTAGAGTTCTTGTGTGTTCTTTTATT
GTTTTTCACTTGATCTTGTGTCATCTCATTTACTTATAATTGACTGTCCTTATAGTTG
ATGTTTTTAATCTTGTAAAGGTTTAAACCCAAATTAATAATCTTTTGAACAT
GACTGTAAAGAAATTTATATCATAGGCATCACTGGAAGTTAATTTATTTAGATCATCCA
GACTTGAACATATCGACCCAGTGGTTTTAAACATTGTGATGCAGCAGCTGCAACTTAC
ATACTAGAGGAGGAAAGCAGGAGCTGACAAGTCAACTCTAAGGTACAGGATTATTTA
AATATCCATTTATTTCAAATACTACCTTAATGATTTCAATTTCAAGTTTAAAAATGGAG
ACAGAACTTTGGCTTTTTTTTCTTTCTTAGCACTTATCTAGAAGACTGTAAATTTGA
CAGAGAGCGAATAGAACTGTTTGCACGAATATCAGGTTACTTACTTTAAATTTTAA
TTAACAGTACTATTTTTCTTTTACTTTGTTTGTAAAGATGTTTTTTTCTTTCTTTCTT
```

Ensembl genome browser 56: H.sapiens - Transcript summary - Transcript: BMI1-001 (ENST00000376663) - Windows Internet Explorer

http://www.ensembl.org/Homo\_sapiens/Transcript/Summary?db=core;g=ENSG00000168283;r=10:22605381-22620413;t=ENST00000376663

File Edit View Favorites Tools Help

Ensembl genome browser 56: H.sapiens - Transcript s...

Home > Human [GRCh37]

Location: 10:22,605,381-22,620,413 Gene: BMI1 Transcript: BMI1-001

**Transcript-based displays**

- Transcript summary
- Supporting evidence (80)
- Sequence
  - Exons (10)
  - cDNA
  - Protein
- External References
  - General identifiers (4)
  - Oligo probes (20)
  - Gene ontology (10)
- Genetic Variation
  - Population comparisc
  - Comparison image
- Protein Information
  - Protein summary
  - Domains & features (
  - Variations (2)
- External Data
  - Personal annotation
- ID History
  - Transcript his
  - Protein histor

Configure this page  
Manage your data  
Export data  
Bookmark this page

**Transcript: BMI1-001 (ENST00000376663)**

Polycarb complex protein BMI-1 (Polycomb group RING finger protein 4)(RING finger protein 51) [Source:UniProtKB/Swiss-Prot;Acc:P35226]

**Location** [Chromosome 10: 22,610,140-22,620,413](#) forward strand.

**Gene** This transcript is a product of gene [ENSG00000168283](#) - There are 11 transcripts in this gene: [hide transcripts](#)

Name	Transcript ID	Protein ID	Description
BMI1-001	<a href="#">ENST00000376663</a>	<a href="#">ENSP00000365851</a>	protein_coding

**VERY IMPORTANT, EVERYWHERE IN ENSEMBL**

**DATA DISPLAY CAN BE FULLY CONFIGURED**

Transcript summary [help](#) [Supporting evidence »](#)

10.27 Kb Forward strand

Export Image

**Statistics**

**CCDS** Exons: 10 Transcript length: 3,428 bps Translation length: 326 residues

**Type** This transcript is a member of the Human CCDS set: [CCDS7138](#)

**Prediction Method** Known protein coding

Transcript where the Ensembl genebuild transcript and the [Vega](#) manual annotation have the same sequence, for every base pair. See [article](#).

**Alternative transcripts** This Ensembl/Havana merge gene entry corresponds to the following database identifiers:

# Position-independent gene identification

## [3. The “candidate gene” approach]

- The choice of candidates is based on:
  1. Tissue specificity and abundance
  2. Chromosomal location
  3. Sequence information and length
  4. Similarity to other known disease-causing genes
  5. Other characteristics (gene conservation, etc.)

# Position-independent gene identification

## [3. The “candidate gene” approach]

- Usually, you should already be familiar with the background literature/information of the subject that you are studying
- However, specifically for diseases, you can check for example OMIM (online Mendelian Inheritance in Man)

PubMed Home - Windows Internet Explorer

http://www.ncbi.nlm.nih.gov/sites/entrez

File Edit View Favorites Tools Help

PubMed Home

NCBI

PubMed  
www.pubmed.gov

A service of the U.S. National Library of Medicine and the National Institutes of Health

My NCBI  
Welcome crivolta. [Sign Out](#)

All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books

Search PubMed for muscular dystrophy Go Clear Advanced Search

Preview/Index History Details

me to PubMed

PubMed database comprises more than 15 million citations for biomedical articles from MEDLINE and life science journals. Citations include links to full-text articles from PubMed Central or publisher web sites.

Introducing RRN  
Rapid Research Notes  
Immediate access to preliminary research reports

NCBI H1N1 Flu Resources:

Submit flu sequences to GenBank  
Latest H1N1 citations in PubMed  
MedlinePlus (consumer health information)  
Enviro-Health links

FLU.GOV  
Know what to do about the flu.  
VISIT FLU.GOV  
SHARE THIS WIDGET

Entrez PubMed  
About Entrez  
Text Version  
Entrez PubMed  
Overview  
Help | Feedback  
Tutorials  
New/Noteworthy  
E-Utilities  
PubMed Catalog  
Journals  
MeSH Database  
Single Citation Matcher  
Batch Citation Matcher  
ClinicalTrials.gov  
Specialized LinkOut  
My NCBI  
Related  
Order Depository  
NLM Molecular  
NLM Catalog  
NLM Gateway  
TOXNET  
Consumer Health  
Clinical Alerts  
ClinicalTrials.gov  
PubMed Central

# Position-independent gene identification

## [3. The “candidate gene” approach]

- The choice of candidates is based on:
  1. Tissue specificity and abundance
  2. Chromosomal location
  3. Sequence information and length
  4. Similarity to other known disease-causing genes
  5. Other characteristics (gene conservation, etc.)

Human (Homo sapiens) Genome Browser Gateway - Windows Internet Explorer

http://genome.ucsc.edu/cgi-bin/hgGateway

File Edit View Favorites Tools Help

Human (Homo sapiens) Genome Browser Gateway

Home Genomes Blat Tables Gene Sorter PCR Session FAQ Help

### Human (*Homo sapiens*) Genome Browser Gateway

The UCSC Genome Browser was created by the [Genome Bioinformatics Group of UC Santa Cruz](#).  
Software Copyright (c) The Regents of the University of California. All rights reserved.

clade genome assembly position or search term image width

Mammal Human May 2004 chrX:24,305,351-24,305,935 1000 submit

[Click here to reset](#) the browser user interface settings to their defaults.

add custom tracks configure tracks and display clear position

#### About the Human May 2004 (hg17) assembly ([sequences](#))

The May 2004 human reference sequence (NCBI Build 35) was produced by the International Human Genome Sequencing Consortium.

#### Sample position queries


A genome position can be specified by the accession number of a sequenced genomic clone, an mRNA or EST or STS marker, or a cytological band, a chromosomal coordinate range, or keywords from the GenBank description of an mRNA. The following list shows examples of valid position queries for the human genome. See the [User's Guide](#) for more information.

**Request:**

chr7  
20p13  
chr3:1-1000000  
chr3:1000000+2000  
D16S3046

Displays a region of chr3 that spans 2000 bases, starting with position 1000000

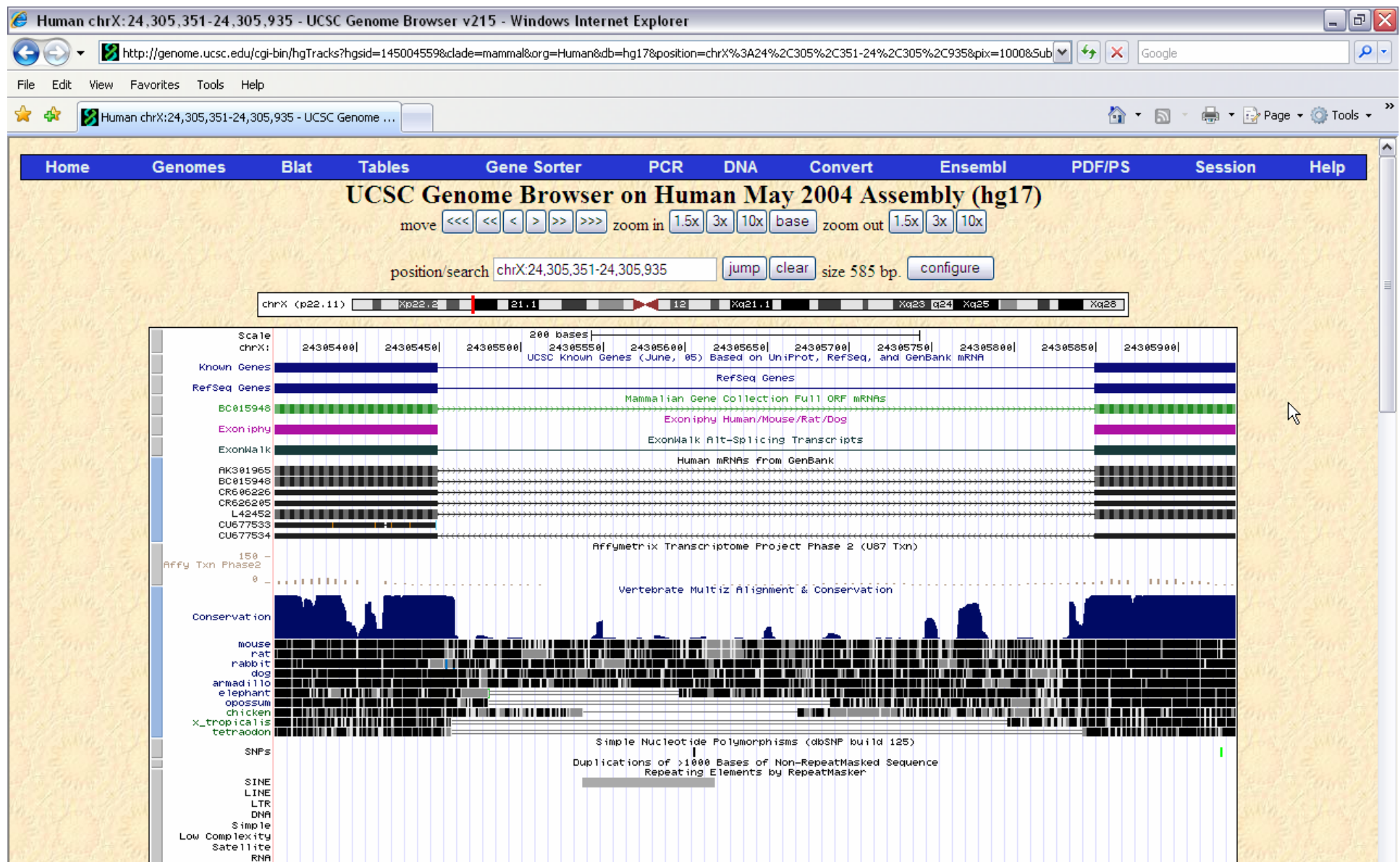
Displays region around STS marker D16S3046 from the Genethon/Marshfield maps. Includes 100,000 bases on each side as well.



*Homo sapiens*  
(Graphic courtesy of [CBSE](#))

# UCSC Genome Browser





collapse all      Use drop-down controls below and press refresh to alter tracks displayed.  
Tracks with lots of items will automatically be displayed in more compact modes.      expand all

**Mapping and Sequencing Tracks**      refresh

<a href="#">Base Position</a> dense ▾	<a href="#">Chromosome Band</a> hide ▾	<a href="#">STS Markers</a> hide ▾	<a href="#">FISH Clones</a> hide ▾	<a href="#">Recomb Rate</a> hide ▾	<a href="#">Map Contigs</a> hide ▾
<a href="#">Assembly</a> hide ▾	<a href="#">Gap</a> hide ▾	<a href="#">Coverage</a> hide ▾	<a href="#">BAC End Pairs</a> hide ▾	<a href="#">Fosmid End Pairs</a> hide ▾	<a href="#">GC Percent</a> hide ▾
<a href="#">WSSD</a> hide ▾	<a href="#">Short Match</a> hide ▾	<a href="#">Restr Enzymes</a> hide ▾			

**Phenotype and Disease Associations**      refresh

[RGD QTL](#)  
pack ▾

**Genes and Gene Prediction Tracks**      refresh

<a href="#">Known Genes</a> dense ▾	<a href="#">CCDS</a> hide ▾	<a href="#">RefSeq Genes</a> dense ▾	<a href="#">Other RefSeq</a> hide ▾	<a href="#">MGC Genes</a> pack ▾	<a href="#">ORFeome Clones</a> hide ▾
<a href="#">Vega Genes</a> hide ▾	<a href="#">Vega Pseudogenes</a> hide ▾	<a href="#">Ensembl Genes</a> hide ▾	<a href="#">AceView Genes</a> hide ▾	<a href="#">ECgene Genes</a> hide ▾	<a href="#">N-SCAN</a> hide ▾
<a href="#">SGP Genes</a> hide ▾	<a href="#">Geneid Genes</a> hide ▾	<a href="#">Genscan Genes</a> hide ▾	<a href="#">Exoniphy</a> dense ▾	<a href="#">Augustus</a> hide ▾	<a href="#">Yale Pseudo</a> hide ▾
<a href="#">ACEScan</a> hide ▾	<a href="#">EvoFold</a> hide ▾	<a href="#">sno/miRNA</a> hide ▾	<a href="#">ExonWalk</a> dense ▾		

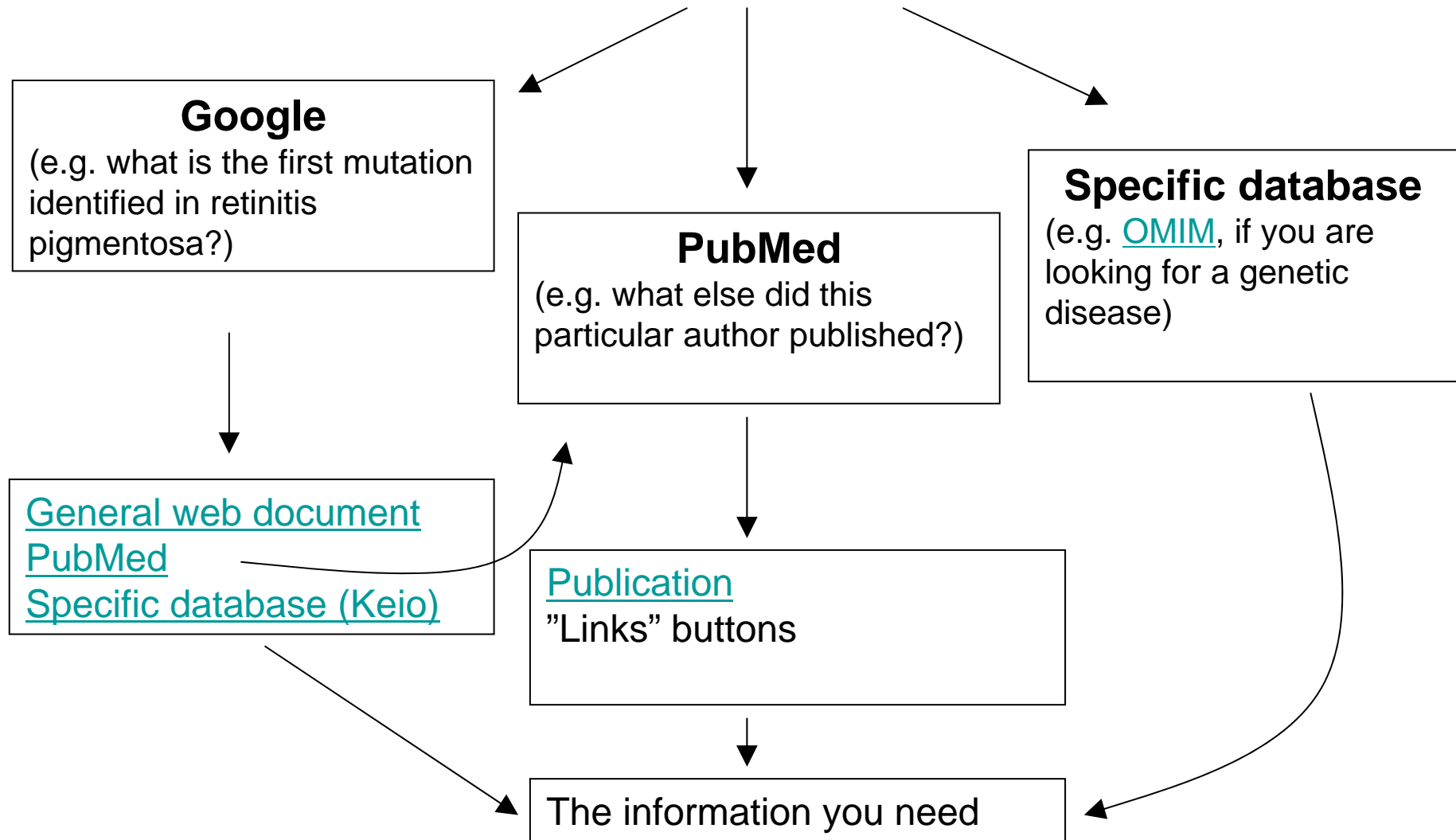
**mRNA and EST Tracks**      refresh

<a href="#">Human mRNAs</a>	<a href="#">Spliced ESTs</a>	<a href="#">Human ESTs</a>	<a href="#">Other mRNAs</a>	<a href="#">Other ESTs</a>	<a href="#">H-Inv</a>
-----------------------------	------------------------------	----------------------------	-----------------------------	----------------------------	-----------------------

# Data integration

## Biomart/Martview

## General, not necessarily gene-related question



**Gene-related question (I am looking for this specific gene's feature)**

**BLAST**

(only if all you have is a sequence)

**NCBI ENTREZ portal**

A little bit less about the gene, but much of anything else

**NCBI's EntrezGene**

More complete and with the "Links" link

**ensembl.org**

Everything about the gene

**GeneCards**

Everything about the gene  
(in a different way...)

Higher Eukaryotes (animals)

**Gene-related question (I am looking for this specific gene's feature)**

**G O O G L E**

**BLAST**  
(only if all you have is a sequence)

**Get gene name**

**General microbial db**  
e.g. [cmr.tigr.org](http://cmr.tigr.org)

**General yeast db, e.g.**  
[genome-www.stanford.edu](http://genome-www.stanford.edu)

**General plant db, e.g.**  
[UGeorgia plant genome page](http://UGeorgia.plant.genome.page)

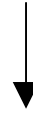
Other organisms

**Specific microbial db**  
e.g. subtilist

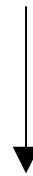
**Specific yeast db**  
e.g. yeastgenome

**Specific plant db**  
e.g. maizegdb

**Question about genome region or group of genes**



**Higher eukaryotes (animals)**



**GO TO GENOME  
BROWSERS**

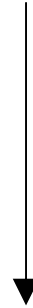
[ENSEMBL](#) (browser)

[NCBI MapViewer](#)

[UCSC genome browser](#)

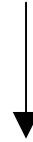


**Question requireing the integration of several databases**



**Biomart/Martview**

**Other genetic characteristics (markers, expression, etc.)  
or specific bioinformatics tools (translation, rev. compl., etc.)**



**Specific web sites**

Expasy tools  
NCBI tools  
ClustalW  
Genecards  
Gene Ontology  
HapMap  
...

## BIOINFORMATIC TOOLS

## A brief summary of some common “techniques”

- The bioinformatic tool to perform the operation you want has probably been already developed: just find Google or NCBI tools (or similar web portals) to find it. (Ex. Reverse-Complement)
- Most genes are already annotated: go directly to databases where you can find the most abundant information on what you are looking for (Ex. Ensembl for gene structures, Swissprot for proteins, etc.)
- Don't be afraid to use Word or other text editors to do your sequence searches. These programs allow to perform Find+Replace operations very easily (Ex. Carriage return at the end of sequences:

For example, looking for “GGCAAAA”

here

```
ATTGTGACCAGGATGCA  
AAAAAA
```

or here

```
ATTGTGACCAGGATGGCAAAAAA
```

is different

- Use the “keyboard BLAST” CTRL + F (or Apple +F) as frequently as you can